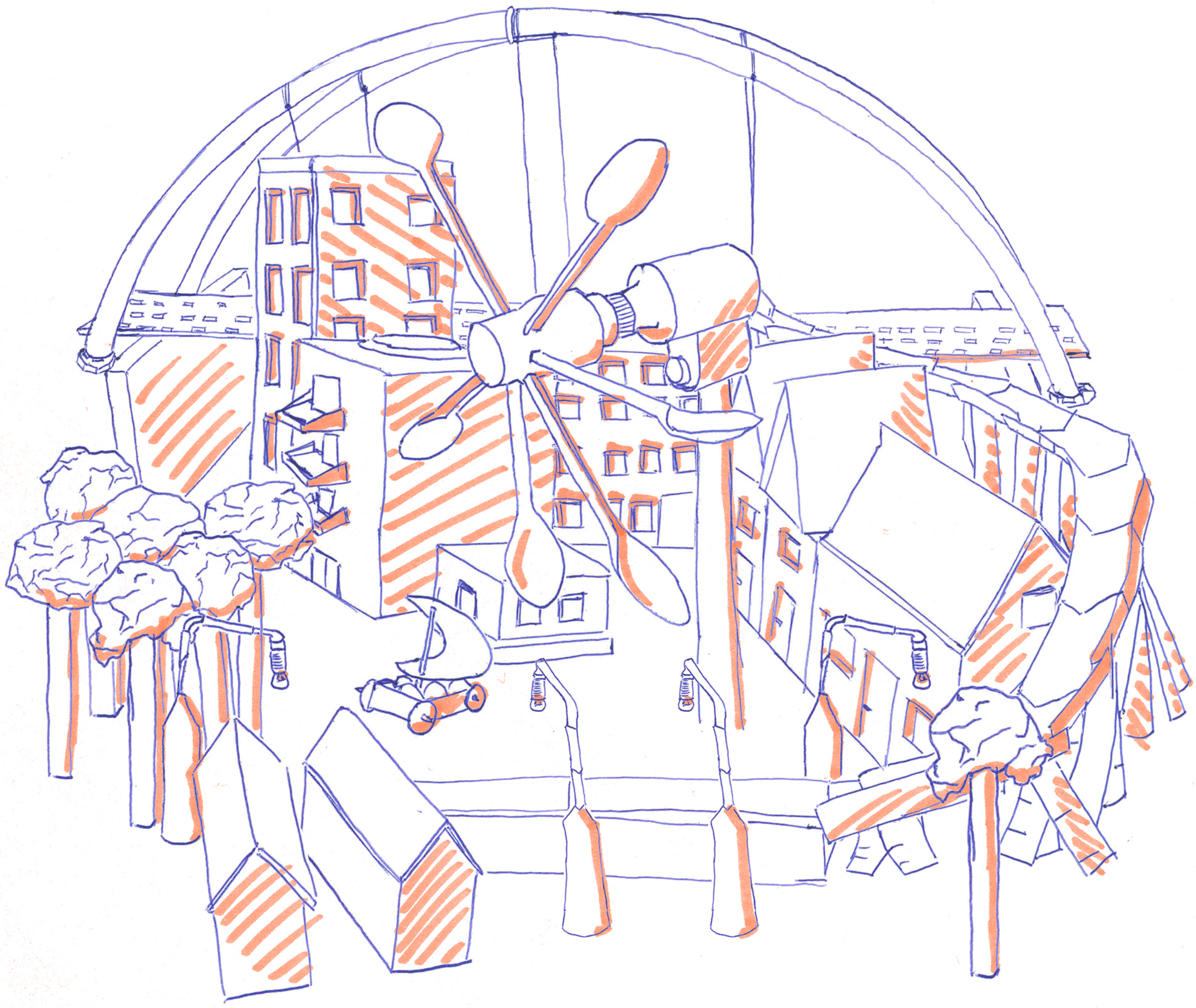
FACULTEIT PSYCHOLOGIE EN
PEDAGOGISCHE WETENSCHAPPEN
CENTRUM VOOR ERVARINGSGERICHT ONDERWIJS

OPEN-ENDED SCIENCE AND TECHNOLOGY LEARNING ENVIRONMENTS.

CHALLENGES FOR PUPILS, TEACHERS AND RESEARCHERS.

MIRANDA THYS



Promotor
Prof. dr. Ferre Laevers

Co-promotoren:
Prof. dr. Lieven Verschaffel
Prof. dr. Wim Van Dooren

Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Pedagogische Wetenschappen

2016

Samenvatting

De behoefte aan krachtige leeromgevingen en professionele initiatieven om leerlingen en leerkrachten van de lagere school te betrekken in W&T, wordt in de literatuur algemeen erkend. Leeromgevingen waarin leerlingen een hoge mate van initiatief krijgen, onder andere project-gebaseerde leeromgevingen (Barak & Raz, 2000; Barak & Doppelt, 2000), kunnen bijdragen tot leerlingbetrokkenheid (Wurdinger, Haar, Hugg, & Bezon, 2007). Ondanks de veelbelovende effecten van deze 'open' W&T leeromgevingen, zijn de condities waaronder ze effectief zijn, nog niet grondig onderzocht. Het kan niet verondersteld worden dat leerkrachten en leerlingen onmiddellijk hun nieuwe rol vinden bij het implementeren van dergelijke omgevingen. Deze doctoraatsverhandeling beoogde de doorslaggevende factoren voor een succesvolle implementatie van zo'n leeromgeving te ontrafelen.

De doelen van dit doctoraatsonderzoek waren tweevoudig. Ten eerste hadden Studies 1 en 4 tot doel de bestaande literatuur wat betreft instrumenten voor de evaluatie van de kwaliteit van (project-gebaseerde) W&T leeromgevingen, te onderzoeken, evenals bij te dragen tot instrumentontwikkeling. Studie 1, een reviewstudie aangaande instrumenten in het veld, toonde het bestaan van een verscheidenheid aan schalen, items en vragen. De resultaten helpen onderzoekers bij het maken van een keuze wanneer zij bepaalde aspecten van de W&T leeromgeving beogen te evalueren, en stimuleren tot het samenstellen van nieuwe instrumenten. In Studies 2 en 3 zijn we zelfs een stap verder gegaan. Omdat in het diepgaande literatuuronderzoek naar instrumenten geen tool werd gevonden die de rol van de leerkracht integraal meet, werd het Classroom Assessment Scoring System (CLASS) (Pianta et al., 2012), een kwaliteitsvol instrument dat ook in andere onderwijsdomeinen gebruikt wordt, geselecteerd. Na inzetten van het instrument in Studies 2 en 3 ontstond de mogelijkheid om de kwaliteiten van de Adult Style Observation Schedule (ASOS) (Laevers & Heylen, 2013), een tool ontwikkeld aan het Centrum voor Ervaringsgericht Onderwijs om leerkrachtstijl te onderzoeken (Studie 4). De resultaten van de convergente validiteitsanalyses toonde aan dat de verwachtte relaties tussen de CLASS en de ASOS dimensies, niet eenduidig gevonden werden. Meer bepaald toonde Stimulerend Tussenkomen (ASOS) congruentie met Content Understanding (CLASS), en Gevoeligheid voor Beleving (ASOS) correleerde positief met Positief Klimaat (CLASS). Er werd geen evidentie gevonden voor de convergentie van Autonomie Verlenen (ASOS) met gelijkaardige CLASS dimensies.

Ten tweede werden in studies 2 en 3 de effectiviteit van het implementeren van een open W&T leeromgeving, het project Dorp Op School (DOS), en factoren gerelateerd aan deze effectiviteit en de implementatie, onderzocht. In Studie 2 werd de evolutie in leerlingbetrokkenheid nagegaan en mogelijke verklarende factoren in het competentieprofiel van leerkrachten – attitudes ten aanzien van W&T (en lesgeven in W&T) en leerkrachtstijl – voor de verschillen tussen scholen en/of klassen in deze evolutie werden geëxploreerd. De hoofdconclusies wijzen erop dat (a) leerlingen groeiden in hun betrokkenheid doorheen het traject en (b) de groei in betrokkenheid (post-pre) positief samenhing met Leerkracht Sensitiviteit, maar negatief met Positief Klimaat en Diepgaand Begrijpen bij controleren voor andere CLASS dimensies. In Studie 3 werd niet alleen de wijze waarop het competentieprofiel voor het project de leerkrachtstijl tijdens DOS bepaalde, maar ook de evolutie van leerkachten tijdens en na de interventie, onderzocht. Het niveau van Emotionele Ondersteuning tijdens DOS hing positief samen met de initiële Emotionale Ondersteuning bij controleren voor de attitudes en de andere CLASS dimensies, maar negatief met Klas Organisatie/attitude t.a.v. onderzoekend leren. Leerkrachten groeiden niet in hun attitudes, maar de interacties wat betreft Oog voor Leerling Perspectieven en Kwaliteit van Feedback verbeterden. Verrassend was de groei in Negatief Klimaat. De resultaten van Studies 2 en 3 werden bediscussieerd in het licht van de uitdagingen die gepaard gaan met de implementatie van DOS.

Summary

In the literature, the need for powerful learning environments and professional initiatives to engage primary school pupils and teachers in science and technology (S&T) is widely recognised. Learning environments in which students are given high levels of autonomy, among them project-based learning environments (Barak & Raz, 2000; Barak & Doppelt, 2000), can contribute to pupils' engagement (Wurdinger, Haar, Hugg, & Bezon, 2007). Despite the promising effects of these 'open-ended' S&T learning environments, until now the conditions for their effectiveness were not thoroughly investigated. It cannot be assumed that teachers and pupils immediately find their new role when implementing such learning environments. This doctoral dissertation disentangles the decisive factors for a successful implementation of such an environment.

The aims of this doctoral research were twofold. Firstly, Studies 1 and 4 aimed to investigate the existing literature on instruments to assess the quality of (project-based) S&T learning environments, as well as to contribute to instrument development. Study 1, a review study with regard to the instruments in the field, revealed that a variety of scales, items and questions. The results help researchers in the field in choosing the best suited instrument to evaluate particular aspects of the S&T learning environment, and stimulates them to compose new instruments. In the intervention research as conducted in Studies 2 and 3, we even went a step further. As no instrument that comprehensively evaluates the teacher's role was found in the profound literature search, the Classroom Assessment Scoring System (CLASS) (Pianta et al., 2012), a high-quality observation tool also used in other fields of education, was selected. After conducting Studies 2 and 3 the opportunity arose to explore the qualities of an existing instrument developed at the Centre for Experiential Education, the Adult Style Observation Schedule (ASOS) (Laevers & Heylen, 2013). The results of the concurrent validity analyses showed that the expected relationships of the CLASS dimensions with the ASOS dimensions, were not univocally found. Stimulation (ASOS) showed congruence with Content Understanding (CLASS), and Sensitivity (ASOS) correlated positively with Positive Climate (CLASS). No evidence emerged for the convergence of Giving Autonomy (ASOS) with similar CLASS dimensions.

Secondly, in Studies 2 and 3 the effectiveness of the implementation of an open-ended S&T learning environment, the Village@School project, and factors related to this effectiveness and implementation, were investigated. In Study 2, pupils' evolution in engagement was studied and possible explaining factors in teachers' competence profile – their attitudes towards S&T (teaching) and their teacher style – for the differences between schools and/or classes with regard to this evolution were explored. The main findings of this study indicated that (a) pupils grew in their engagement throughout the trajectory and (b) the growth in engagement (post-pre) was positively related to Teacher Sensitivity, but negatively to Positive Climate and Content Understanding when controlling for other CLASS dimensions. In Study 3 not only the way in which teachers' competence profile before the project determined their teacher style during Village@School, but also the evolution in this profile throughout and after the intervention, was explored. The level of Emotional Support during Village@School related positively to the initial level of Emotional Support before the project when controlling for teachers' attitudes and the other CLASS domains, but negatively with Classroom Organisation/Attitude towards inquiry learning and Emotional Support during the project. Teachers didn't grow in their attitudes, but the interactions involving Regard for Student Perspectives and Quality of Feedback improved. Surprisingly, a growth in Negative Climate was detected. The findings of Studies 2 and 3 were discussed in the light of the challenges involved in the implementation of the Village@School project.

Table of contents

General introduction

Sketching the problem

Nowadays in society, there is a high demand for scientists as not many students[1] choose a STEM (science, technology, engineering, mathematics) career (Organisation for Economic Co-operation and Development, 2007). Research has shown that this is related to negative attitudes in students towards the disciplines of science (often called 'inquiry') and technology (often referred to as 'design'), and these are formed at a young age (Blatchford, 1992; Murphy & Beggs, 2003; Pell & Jarvis, 2001). The seeds of negative attitudes towards science and technology (S&T[2]) are even sewn in primary school. Pell and Jarvis (2001), carried out research with 800 pupils in English primary schools and found that pupils' interest in science fell as they moved up through primary school. Younger pupils (8-9 years) have considerably more positive attitudes towards science than older pupils (10-11 years) (Murphy & Beggs, 2003). Either no or not much education in these fields is provided to pupils, or the education which is provided is not engaging for them. Related to this, a lot of primary school teachers, also in Flanders and The Netherlands, often have negative attitudes towards S&T and are not yet specifically educated to teach these disciplines (Cobern & Loving, 2002). This results in much insecurity in teaching S&T.

[1] In the literature the concept of 'students' is often used in different ways; not only to point to primary but also to secondary school learners. In what follows, normally this term will be used when discussing the literature; the term 'pupils' will be only used when authors themselves use the concept to refer to primary school learners. As the own research was conducted in primary school, the term 'pupils' is used.

[2] Although 'S&T' will be discussed here, more studies have been conducted which only related to science, whether or not containing technology. In order to provide clarity when giving an overview of the literature, the concept as used by the authors of the different studies will be used.

Taken together, there is a need for powerful learning environments and professional development initiatives in S&T education to engage pupils and teachers in (the learning and teaching of) S&T. When it comes to high-quality S&T education, the literature points to environments in which students are given high levels of autonomy, in the context of self-regulated (Boekaerts, 2002), project- (Barak & Doppelt, 2000; Barak & Raz, 2000), problem- (Barron, Schwartz, Vye et al., 1998; Savery & Duffy, 1995), inquiry- (Furtak, Seidel, Iverson, & Briggs, 2012) and design-based learning environments (Fortus, Dershimer, Krajcik, Marx & Mamlok-Naaman, 2004). Due to their student-centeredness, from now on they will be referred to as 'open-ended' learning environments. These learning environments have led to different positive outcomes in students, not only with regard to their learning of S&T, but also with regard to their motivation (Barak & Raz, 2000; Barron et al., 1998; Doppelt, 2003; Liu & Hsiao, 2002; Westwood, 2006, in Kaldi et al., 2010) and engagement (Kaldi, Filippatou, & Govaris, 2011; Wurdinger, Haar, Hugg, & Bezon, 2007).

However, most research on open-ended S&T learning environments has been conducted in rich demonstration sites or in classes taught by researchers, thereby limiting our understanding of how inquiry teaching and learning look in an ordinary classroom taught by teachers (Krajcik, Blumenfeld, Marx et al., 1998). Moreover, because of their innovative character this type of learning environment may be uncomfortable for teachers and pupils, as they are not used to it. Project-based science teaching and learning involve complex role changes for teachers and students (Polman & Pea, 2001). Hands-on inquiry redistributes the responsibility of learning to the student (Osborne, 1996; White, 1988, both in Zion & Slezak, 2005), and the student's role changes from a passive recipient of information to a constructivist participant in the creation of understanding (Driver et al., 1994; Rossman, 1993; Wheatley, 1991). We cannot assume that there is a straightforward positive relation between implementing such learning environments and the teachers and students immediately finding their new role. This doctoral dissertation was born out of a need to disentangle the decisive factors for a successful implementation of such a learning environment in the daily Flemish[3] and Dutch primary classroom. It is interesting to explore the factors that (1) make pupils grow in the way in which they make use of the opportunities that are provided to them (this is measured via their growth in engagement) and the factors that (2) make teachers grow with regard to their initial way of dealing with S&T activities (as measured via classroom interactions) as well as with regard to their attitudes. The project-based learning environment Village@School, designed and implemented for the first time in 2008 by the Centre for Experiential Education (CEGO), IMEC and the Roger van Overstraeten (RVO)-society (De Winter, Van Cleynenbreugel, Buyse, & Laevers, 2010), functions as the implemented S&T learning environment for the different empirical studies in this dissertation. In what follows, the way in which CEGO provided the context for this doctoral research is described.

[3] Flanders is the northern part of Belgium.

Context of the PhD project

As a result of its earlier studies conducted with regard to S&T learning environments, CEGO became part of Curious Minds, set up in 2011 and financed by the Dutch Platform Bèta Techniek. Curious Minds is a research programme executed by a consortium of seven universities – six Dutch and one Flemish – studying the talents of children (3-13 years old) in combination with S&T learning environments, "excellent" learning contexts which foster those talents (www.talentenkracht.nl). Consisting of 14 different research projects, the goal of the consortium is to further research on scientific and technological talents and excellence and to provide professionals and parents with insights and instruments for talent development. This doctoral research, in combination with another doctoral study also conducted at CEGO (by Veerle De Winter), forms a part of these projects.

In light of the sketched research aims, CEGO, because of its conceptual framework and previous research, was a suitable context to conduct this doctoral research. The experiential approach, in which one takes the perspective of the learner through a reconstruction of his/her experiences, forms the basis of the process-oriented research conducted at the centre (Laevers, 2000). Grounded in this approach, concepts and instruments (the observation tools and questionnaires) have been developed and can be divided into three categories: the process in learners, the learning environment that fosters the process and learners' outcomes.

In the 'process' category, well-being and involvement are seen as key indicators for the power of the learning environment, and are situated between the learning environment and the outcomes. 'Well-being' refers to children feeling at ease, acting spontaneously, and showing vitality and self-confidence (Laevers, 1999). 'Involvement' is characterised by concentration and intense mental activity (Laevers, 2011). This latter key indicator matches the commonly-used concept 'engagement' (Fredricks, Blumenfeld, & Paris, 2004; Lawson & Lawson, 2013) and also the 'state of flow' as conceived by Csikzentmihalyi (1999). The Leuven Involvement Scale (LIS), as an operationalisation of the concept, is being used in a broad range of research to assess involvement in learners – from babies to adults.

Through action research, practice-based evidence has been gathered with regard to the conceptualisation and implementation of the learning environment that fosters and favours well-being and involvement. According to Laevers (2007), the basis is an open framework approach in which teachers and learners have a high level of initiative in co-construction of learning (Laevers, 2006). Seven factors have been identified in this category: among these the richness of the offer of activities and materials, the level of initiative provided to learners, and the provision of opportunities for collaboration. Apart from that, the quality of the interactions between teachers and students is crucial. This quality is captured in the concept of teacher style, which is evaluated by using the Adult Style Observation Schedule (ASOS).

Finally, in the category of the outcomes, the Piaget-inspired concept of 'schema' provides a basis to describe, observe and assess the competences and dispositions in learners. For several developmental domains, concepts, scales and tests have been developed, particularly for self-organisation and entrepreneurship, social competences and understanding of the physical world; the latter entitled 'Minitest – Eye for Science and Technology' [Oog voor Wetenschap & Techniek (OWT)] (De Winter, Van Cleynenbreugel, Buyse, & Laevers, 2010).

Through its research, CEGO aims to explore and understand the relation between the process, the learning environment and the outcomes. This doctoral research will focus on the first and second categories, as the learning environment in itself - particularly the teacher acting in this learning environment - as well as the learning environment in relation to pupils' involvement or engagement, is explored. While this doctoral research does not shed light on the category of the outcomes, this constitutes the main subject of the parallel doctoral research conducted by Veerle De Winter. In the latter piece of PhD research, the potential of Village@School for pupils' and teachers' development of competences with regard to the understanding of the physical world is closely examined.

In the remainder of this introduction, the different concepts that constitute the theoretical framework of this PhD research will be discussed. In the first part, open-ended S&T learning environments and their implementation will be explored. Firstly, there will be a focus on the characteristics of open-ended learning environments and the operationalisation of their implementation. Secondly, literature will be presented with regard to factors in teachers (i.e. their competence profiles) which may affect this implementation. Finally, the effectiveness of the implementation of project-based learning environments, in terms of pupils' growth in engagement and teachers' growth in their initial competence profile, will be discussed. The second part of this general introduction provides the specific research questions, how they will be reached in the different studies and the general research design.

Theoretical framework

Open-ended S&T learning environments

Open-ended S&T learning environments, in which students are provided with high levels of autonomy, typically belong to the constructivist tradition of education (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Fallik, Eylon, & Rosenfeld, 2008; Piaget & Inhelder, 1969). In the following overview, the main focus will rest on the literature on project-based learning environments, but the characteristics that will be described are similar to those of inquiry- and design- based learning environments. In fact, in project-based learning environments S&T can, but do not have to, come together (Barak & Raz, 2000). Science, and particularly doing science as inquiry, refers to engaging students in scientifically-oriented questions, finding evidence and coming up with explanations (National Research Council, 1996). Technology, and particularly technology as design, does not refer to the use of digital or information technology, but to the process of 'designing' technological solutions (NDET, 2006, in Hansen, 2010; Roth, 2001). Technology is the product that includes "all the knowledge that its creation entails, through which people utilise their resources on earth so that they can first survive and then move on to raise their standard of living" (Hansen, 1997, p.112).

There is a diversity of features that define project-based learning environments, and therefore there is no universally accepted model or theory of project-based learning environments (Thomas, 2000). In particular, a high level of initiative provided to learners (Thomas, 2000, p. 285), the provision of an authentic task (e.g. Hmelo-Silver, 2004) and intensive collaboration among learners (Kaldi et al., 2011) are highlighted. Although initiative to learners is provided, the teacher still has an active role and acts as a tutor, a guide and a partner in the learning process (Barth, 1972). However, there is currently no general agreement on the characteristics of such a coaching role in these learning environments (Hakkarainen, 2009; Kolodner, 2001).

While these characteristics are typical for open-ended, and in particular project-based learning environments, the extent to which these characteristics are found in practice will be different from class to class. This mainly has to do with the fact that it is not possible, and even not expedient, to mark the contours of project-based learning environments beforehand as their starting point is the high level of initiative that is given to students. The curriculum is 'emergent' as the teacher's initiative, an offered activity, is responded to by the students and pushed in a certain direction, after which the teacher makes a new contribution that provokes new actions (Laevers, 2011).

When using the term 'implementation' of learning environments, different aspects can be taken into consideration. In the Experiential Education framework, next to the infrastructure and equipment and the content of the specific activities, the concept of teacher style is used (Laevers, 2005). It grasps a teacher's individual pattern of the way in which he or she intervenes in a wide variety of situations. Three dimensions are discerned: stimulation, sensitivity and giving autonomy (Laevers & Heylen, 2013). Teacher style is embedded in the interactions between the teacher and their pupils. Next to these teacher-pupil interactions, collaboration, and so interactions among pupils play an important role in project-based learning environments. It is noteworthy that student-student interactions cannot be seen disconnected from teacher-student interactions; via his or her interactions, the teacher also determines interactions among students (Pianta et al., 2012).

One can say that these interactions are situated at the micro-level, which has to be placed in a context in which the infrastructure and equipment, organisational forms and the content of the specific activities create the necessary conditions for learning. They are part of the construct of 'project-based' learning environments. It is interesting to note that teacher style has a particular status within these broader conditions. For instance, a teacher can provide a high level of initiative by letting pupils autonomously work on a task (macro-level), but on the basis of his or her interactions with learners it becomes clear that he or she gives little responsibility while they are solving the problem (micro-level). To begin with, the broader conditions will be discussed, before focusing on interactions.

A first broader aspect of the learning environment is the amount of initiative given to the pupils in the project or activities provided. Over the course of a longer time or project, one can ask whether the pupils get the opportunity to be the initiators of each of the activities, whether they are in charge in the distribution of tasks, whether they collect the necessary materials themselves, and so on. Secondly, one can ask whether the learning environment is rich enough in terms of materials (such as wood, cardboard, film, electric wire, plastic tubes and so on) and tools (such as hammers, electric drills etc.) available for the design and construction of technological solutions. Thirdly, the kind of activities can be evaluated by determining the opportunities for challenging design and inquiry activities to take place (for example, collecting information through the Internet, designing experiments, drawing plans, or making constructions) that are meaningful and relevant to children (Alexander & Wade, 2000; Curtis, 2002).

The importance of the micro-level of interactions in this kind of learning environment has previously been described in the literature. Children need ample and high quality opportunities to talk to others, such as peers and teachers, about discoveries, ideas and solutions (Damhuis & De Blauw, 2011). Without such opportunities, S&T will not reach its goals. This is related to the trap of the so-called

'pseudo-inquiry' (Harlen & Léna, 2011): plenty of practical activity, but a lack of involvement of the children in making sense of phenomena or events in the natural world [emphasis in the original]. However, Polman and Pea (2001, in Zion & Slezak, 2005, p. 877) claimed "that additional research is needed to refine and expand our understanding of how to effectively guide students in long-term open and full inquiry tasks".
As discussed earlier, student-student interactions as well as teacher-student interactions take place. In what follows, a close look at each of these interactions will be provided. Firstly, as collaboration is a typical characteristic of open-ended learning environments, interactions with peers are particularly relevant. Students have shared problems, and have to come to a consensus on possible solutions and find a mutually acceptable way to solve them (Liljeström, Enkenberg, & Pöllänen, 2013). This process enhances reflective communication and is the highest form of collaboration (Engeström, 1992). This synergy can foster the clarification of ideas (Kaldi et al., 2011), the internalisation of content knowledge (Cross, 1998), deep learning (Kolodner, 2006a) and positive interdependence (Kaldi et al., 2011).

Secondly, teachers cannot stay in the background in this type of learning environment (Liljeström et al., 2013). The role they have to make it work, is different. They are no longer - as traditionally - supposed to give all pre-determined aims and guidelines; nor do they need to engage students in experiments of which they already know the correct answers (Mueller, 2011). Instead, other ways in which the teacher can interact are promoted. "The teacher must guide, focus, challenge, and encourage student learning" (American Association for the Advancement of Science (AAAS), Project 2061, 1993; NRC, 2000, in Zion & Slezak, 2005, p. 877). However, the literature is not clear about the characteristics of the teacher's role in open-ended S&T learning environments (Polman & Pea, 2001). Several varying characteristics of these interactions are emphasised, but there is still a vagueness with regard to the relative importance of particular ways of interacting with students in such learning environments. In general it is emphasised that interaction in S&T needs to be thought- and talk-provoking (Damhuis & De Blauw, 2011). Interactions should stimulate pupils' mental activity (Laevers, 2011). Firstly, authentic questions – which have no pre-specified answers – and higher-level questions, especially student-generated questions, are particularly important because they are substantively engaging (Nystrand, Wu, Gamoran, Zeiser, & Long, 2003). Secondly, a teacher needs to be able to integrate students' everyday experiences and raise these experiences as resources that can be shared by the whole learning community (Viilo, Seitamaa-Hakkarainen, & Hakkarainen, 2011). Thirdly, deepening feedback and building further on pupils' responses matters (Damhuis & De Blauw, 2011). Fourthly, classroom management (Jacobowitz, 1997; Lawson, 1995) and an efficient use of classroom time remain important in these open-ended settings. Finally, a few authors in the field point to more affective characteristics of the interactions, such as a positive class climate (e.g. Doppelt, 2003).

A number of studies which have investigated interactions in the field of S&T education have only focused on classroom discourse (Chin, 2007; Erdogan & Campbell, 2008; Hackling, Smith, & Murcia, 2011; Reinsvold & Cochran, 2012; Scott, Mortimer, & Aguiar, 2006; Smart & Marshall, 2013), or the dialogues between teachers and students and among students (e.g. Pehmer, Gröschner, & Seidel, 2015). More specifically, in these studies the written and spoken language of these dialogues is analysed (Mercer, 2010; Nystrand et al., 2003). The nature of the questions asked in the classroom – both by teachers and pupils – is studied, and teachers' elaboration of pupils' answers is evaluated. As such, one can study whether questions elicit and scaffold students' ideas (Smith, Blakeslee, & Anderson, 1993) and to what extent the teacher keeps a leading role in classroom discussions. While the study of this classroom discourse mainly deals with the evaluation of the extent to which the teacher shows authoritativeness and permissiveness in his or her interactions with learners, interactions also have other aspects which are worth investigating. In a broader perspective on interactions, as suggested and proven by researchers who conduct research in other fields of education (e.g. Pianta, Hamre, & Mintz, 2012), the quality of the relations in class, the teacher's sensitivity, his or her management of student behaviour and classroom activities, whether teachers and students build on (other) students' responses, and so on are also taken into consideration when studying interactions in the classroom.

However, on the basis of the literature on open-ended S&T learning environments, one can conclude that there is a lack of instruments that comprehensively measure this quality of interactions, or the quality of S&T learning environments in general. This may also explain why our insight into the effective aspects of open-ended S&T learning environments is limited (Polman & Pea, 2001). Investigating these aspects more closely may not only contribute to the research in the field, but may also provide primary school teachers with clearer information about their roles in such learning environments; something which is necessary to develop their competences in using this kind of learning environment in their classrooms. A good starting point may be to develop new instruments on the basis of the existing scales. Other from that, one could test instruments which have been used in other fields of education, but have a theoretically similar connection to the literature relating to S&T learning environments.

Pupils' engagement (growth)

Pupils' engagement may have different conceptualisations, but researchers often use a combination of three dimensions in their conceptualisation of engagement (Fredricks et al., 2004). The emotional dimension deals with showing interest, giving value and positive affect; the cognitive dimension with the use of cognitive and metacognitive strategies; and the behavioural dimension with showing effort, persistence and seeking help (Linnenbrink & Pintrich, 2003). At CEGO the concept of 'involvement' was developed to determine whether the learning environment leads to the outcome (Laevers & Heylen, 2003). Involvement is characterised by concentration and intense, intrinsically motivated mental activity (Laevers, 2011). It connects with the 'state of flow' as conceived by Csikszentmihalyi (1988). Although the three dimensions of engagement (Fredricks et al., 2004; Lawson & Lawson, 2013) can be recognised in the concept of involvement, the CEGO approach of this phenomenon is holistic in nature. As a consequence in the assessment of levels of involvement only one overall score is attributed. In what follows, the concept of 'engagement' will be used, as this term is more commonly applied in the literature.

Students' engagement in open-ended S&T learning environments has previously been investigated quite often (e.g. Cornell & Clarke, 1999; Mant, Wilson, & Coates, 2007). The main conclusion of these studies is that students are more engaged when involved in such learning environments than in teacher-directed settings. The positive results are attributed to their typical characteristics (hands-on activities, working together, more thinking for themselves, having autonomy, and so on.). However, studies also show that there is no guarantee of high engagement in students in these learning environments, as students may be affectively engaged but not necessarily cognitively engaged. Pupils do not always discuss ideas or use evidence systematically (Germann & Aram, 1996; Palinscar, Anderson, & David, 1993). Therefore, educational researchers in the field (Keys & Bryan, 2001; Krajcik et al., 1998) insist on the investigation of how teachers can make these environments successful for students' engagement and learning via their actions in S&T learning environments. Moreover, not only these actions but also teachers' attitudes towards S&T (and its teaching) may play a role in the determination of pupils' engagement. Until now different studies have investigated the effects of teachers' attitudes towards S&T (teaching) on pupils' attitudes towards the domains of S&T (Harlen & Holroyd, 1997; Jarvis & Pell, 2005), but not on pupils' engagement during S&T activities.

Alongside pupils' engagement, it is also interesting to investigate the possible growth in their engagement, as learning in such environments is often new to them and pupils have to get used to the new way of working. Self-directed learning requires the planning and management of learning by individuals (Skager, 1984), and not just doing whatever one wants to do (Pirozzo, 1987).

Teachers' competence profile in relation to the implementation of open-ended S&T learning environments

By now it may have become apparent that the implementation of project-based learning environments is a major operation, as teachers often face a challenge in implementing the project method (Marx, Blumenfeld, Krajcik, & Soloway, 1997). A successful implementation of project-based learning environments depends on teacher characteristics (such as attitudes, knowledge and skills), integrated into the teacher's competence profile (Mulder, 2001). Particularly in the field of S&T education, many studies have demonstrated a relation between the teacher's attitudes, knowledge, and skills with regard to (the teaching of) S&T on the one hand, and his or her actual teaching of S&T activities on the other (Rohaan, Taconis, & Jochems, 2010). Often, three categories are distinguished in the 'competence' concept: Subject Matter Knowledge (SMK) elements, Pedagogical Content Knowledge (PCK) elements, and Attitude elements (Alake-Tuenter et al., 2012). Although these three categories are often distinguished, Depaepe, Verschaffel and Kelchtermans (2013) show in their review that some authors argue that PCK should not be seen as a separate knowledge base (e.g. Ball, Thames, & Phelps, 2008; Bromme, 1995) that can be distinguished from SMK or from the actual teaching of a subject in a particular context. With this noted, the different concepts that have been used to conceptualise teachers' competence profile will be discussed in what follows.

Firstly, SMK is knowledge about the content that is to be taught. SMK requires knowledge and understanding of facts and constructs of a discipline, as well as the connections between facts and constructs (Alake-Tuenter, Biemans, Tobi, & Mulder, 2013). This specific type of knowledge is necessary to teach science successfully (Katz, Sadler, & Craig, 2005; Lee, Hart, Cuevas, & Enders, 2004). Teachers with a high level of SMK tend to focus more on their students' understanding of the subject matter (Dietz & Davis, 2009).

SMK is necessary but not sufficient for effective teaching. Teachers also need PCK (Shulman, 1986), or knowledge that blends subject matter and pedagogy (Avraamidou & Zembal-Saul, 2010; Davis, 2005). PCK is conceptualised in different ways by various researchers (Depaepe et al., 2013; Van Driel, Beijaard, & Verloop, 2001). In the context of science education, PCK can include knowledge of students' thinking about science, science curriculum, science-specific instructional strategies, assessment of pupils' science learning, and orientations to teaching science (Magnusson, Krajcik, & Borko, 1999; Park & Oliver, 2008). Recently, several researchers have studied the relation of PCK (in terms of quantity and quality) to teacher practice (Loughran, Mulhall, & Berry, 2008; Nilsson & Loughran, 2012). They concluded that this type of knowledge guides a teacher's instructional decisions.

Finally, one's attitude represents "a summary evaluation of a psychological object captured in attribute dimensions as good-bad, harmful-beneficial, pleasant-unpleasant, and likable-dislikable" (Ajzen, 2001, p. 28). Authors such as Van Aalderen-Smeets and Walma van der Molen (2015) make a distinction between personal and professional attitudes towards science. They assume, based on the Theory of Planned Behavior (Ajzen, 2002; Ajzen & Fishbein, 1980), that the professional attitude towards teaching science has the most direct influence on the intention to teach science and to the actual teaching of science. However according to them, this requires that they change as a person, as their personal attitude is a part of who they are. As such, they presume personal attitude towards science to be related to professional attitude. Different authors (Breckler, 1984; Eagly & Chaiken, 1993; Katz & Stotland, 1959; Klop & Severiens, 2007) identified three inseparable components, which can be affective, cognitive and behavioural in nature. Teachers with less positive attitudes towards S&T (teaching) rely more on standardised methods and top-down instruction (Appleton & Kindt, 1999; Harlen & Holroyd, 1997; Jarvis & Pell, 2004; Plonczak, 2008). Particularly in primary education – in which teachers are often not specifically prepared to teach science – this causes a problem (Cobern & Loving, 2002). Therefore, one can suppose that teachers with positive attitudes towards S&T (teaching) are better able to cope with the implementation of innovative S&T learning environments in comparison to others who have less positive or even negative attitudes.

Alongside SMK, PCK and Attitudes, which – however dependent on their conceptualisation – predetermine the actual teacher practice, the teaching practice itself can be examined when studying teachers' competence profiles. In a teacher's practice, their skills become visible. Important teacher skills are revealed in the way in which a teacher interacts with pupils, in his or her own teacher style. With regard to classroom dialogue, researchers found that teachers using rather rigid conversational patterns provide limited opportunities for interactions with the students (Hugener et al., 2009; Jurik, Gröschner, & Seidel, 2013; Lipowsky et al., 2009), and teachers sometimes find it difficult to modify their habitual teacher-student interactions towards a richer and more purposeful verbal exchange (Alexander, 2005; Osborne, Simon, Christodoulou, Howell-Richardson, & Richardson, 2013). However, rich interactions are considered important in order to act as a good coach in open-ended S&T learning environments (Damhuis & De Blauw, 2011).

Teachers' growth in their competence profile

The effectiveness of implementing open-ended S&T learning environments may not only be determined by examining the student's (growth in) engagement and the way in which the teacher implements the project, but also by studying the evolution in teachers' competence profiles. Previous studies have been conducted that investigate the ways in which teachers can grow in their attitudes in S&T learning environments (e.g. van Aalderen-Smeets & van der Molen, 2015). In the end, these positive attitudes may lead to a better teaching of S&T (Osborne & Dillon, 2008; Osborne, Simon, & Collins, 2003). Moreover, previous research showed that the way in which a teacher interacts with his or her students is viable to change; teachers can incorporate high quality interactions in their S&T lessons when they participate in courses especially designed for that purpose (Damhuis & De Blauw, 2011). Different researchers (Damhuis & de Blauw, 2011; Smith, 2015) point to the fact that a single workshop or lecture is not enough. Participation in team meetings and frequent practice in teachers' own classes, (combined with coaching) are also important. However, today, hands-on pedagogical professional development courses for teachers are insufficiently provided (Department of Education and Skills, 2012; Murphy & Smith, 2012; Smith, 2014).

Research goals

The aim of this doctoral research was twofold. Firstly, the research aimed to investigate the existing literature on instruments used to assess the quality of S&T learning environments, particularly project-based S&T learning environments, as well as to contribute to instrument development. Secondly, the effectiveness of the implementation of the project-based S&T learning environment Village@School, and factors related to this effectiveness and the implementation, were investigated in two different ways. On the one hand, pupils' evolution in engagement[4] was investigated, and possible explaining factors in the teacher (their attitudes, and teacher style) for the differences between schools and/or classes with regard to this evolution were explored. On the other hand, the implementation of the project and its effectiveness was investigated by looking at the teacher. At first, the way in which factors in the teacher (their attitudes and style) before the project determined how Village@School was enrolled, in terms of the teacher style, was explored. Next, the teacher's evolution with regard to two aspects of his or her competence profile, i.e. his or her attitudes and his or her teacher style, throughout[5] and after the intervention was investigated. These research goals were reached in the four studies of this dissertation. All studies represent manuscripts that are either published in or have been submitted to international peer-reviewed journals. All manuscripts comprise an abstract, an introduction and theoretical framework, a methodology, a results section, and finally a conclusion and discussion.

In Studies 1 and 4 the first research aim was met. Study 1 consists of a review of existing instruments which measure the quality of S&T learning environments. Based on a close analysis of their operationalization, a conclusion was drawn about the suitability of particular scales, items and questions for project-based S&T learning environments. As the existing offer of instruments is limited, the Classroom Assessment Scoring System (CLASS) which comprehensively measures the teacher style, was selected. As the aim was to gain insight into teacher style in the reality of the teacher's classroom, another criterion for this selection was the opportunity to use the instrument in observations. Therefore, instruments that measure the perceptions of the participants were excluded. In the end, the choice for the CLASS provided us with the opportunity to explore the concurrent validity of the CLASS with the ASOS in Study 4, and by doing so contribute to the development of instruments which measure the quality of the learning environment.

---

[4] As mentioned earlier, in the parallel PhD-study of Veerle De Winter this effectiveness was evaluated by focusing on the effects of the project on pupils' competences and attitudes.
[5] Only for teacher style the growth *throughout* the intervention was investigated.

The second research aim was unfolded in Studies 2 and 3. While in Study 2 the influence of the mean teacher style over the trajectory on pupils' growth in engagement was investigated, in Study 3 the relation between the teacher's initial attitudes and teacher style and his or her teacher style during Village@School was determined. With regard to aspects of the implementation, the intention is not to elaborately analyse the broader conditions of the implementation, by providing insight into the ways in which the project Village@School unfolds; i.e. which activities pupils conduct, and how they design and build particular constructions[6]. Instead, in this dissertation we will dig deeper into the teacher-pupil and pupil-pupil interactions. With regard to the factors preceding the implementation, the teacher's competence profile was used. The teacher's competence profile was operationalised as his or her attitudes towards S&T (and its teaching) and the teacher's style before the project Village@School was initiated in his or her class. While we recognise the importance of a teacher's SMK and PCK, the focus in this dissertation is on the attitudes of the teacher and his or her actual skills during S&T activities[7].

To determine whether the implementation was successful, both pupils' and teachers' evolution was investigated. In Study 2 the possible growth in pupils' engagement[8] was determined and, in case of growth, it was explored whether this growth was attributable to the quality of the interactions and teachers' attitudes towards S&T (and its teaching). Next, in Study 3, the focus was on teachers' competence profile, its influence on the way the project is implemented and the potential evolution of this competence profile. The relationship between this competence profile and the quality of the interactions during the Village@School project was investigated.

[6] Although these broader conditions and their connections with other variables are not the subject of this PhD research, during the intervention research data were collected – but have not yet been analysed – in order to gain insight into these aspects (via teacher interviews and teacher diaries).
[7] In the parallel PhD-study, Veerle De Winter further developed a picture test called 'Minitest – Eye for Science and Technology' [Oog voor Wetenschap & Techniek (OWT)] (De Winter et al., 2010), designed to measure understanding of S&T in pupils and teachers (teachers' SMK).
[8] Analyses were also conducted with pupils' engagement as a dependent variable and in these particular studies, some of these results were reported in the discussion section.

Overall Research Design

Before the actual intervention research began, a pilot study was conducted with 10 teachers within 5 schools in The Netherlands (2 teachers in each school). The main aim of this pilot study was to test the research design and the instruments. In the following description of the general research design of this PhD research, we will refer back to our experiences in the pilot study in order to justify some choices in the final research design.

In the actual intervention research, 34 primary school teachers within eighteen schools volunteered to participate. Some schools and teachers were recruited via the connections of the CEGO in Flanders and the Netherlands, and other school boards were contacted on the basis of the existing list of primary schools in Flanders (www.ond.vlaanderen.be). Four of the schools were located in the Netherlands (8 Dutch teachers) and 14 in Belgium (26 Belgian teachers). The participating teachers taught in the 3rd grade (3 groups of pupils), 4th grade (7 groups of pupils), 5th grade (16 groups of pupils), 6th grade (16 groups of pupils), or in a combination of two consecutive grades. Four classes contained both 5th and 6th grade pupils; three classes had both 4th and 5th grade pupils and three others consisted of both a 3rd and 4th grade pupils. Two of these mixed classes - a 3th and 4th grade and a 5th and 6th grade - belong to a school for highly gifted students. For practical reasons, data was collected in two waves. The first wave (involving 19 teachers) started in November 2013 and ran until June 2014; the second wave (involving 17 teachers) began in September 2014 and lasted until March 2015.

The intervention consisted of the implementation of the Village@School project. Village@School is a project for primary education wherein the development of technological and entrepreneurial competences are the main goals (www.dorpopschool.be). During the project children are given a lot of autonomy while they are challenged to build a miniature site on a standard plate with a budget of 100 euros. To integrate as many technological applications in the design and construction as possible, all conceivable 'problems' relevant from a technological viewpoint need to be 'discovered', solutions have to be found, choices have to be made and constructions need to be built in a context of cooperative learning. In order to clarify the goals of the project to teachers and pupils and set the scene in a more imaginative way, the concept of 'stars' that pupils have to collect was introduced. Each star is connected to the roles which are important to succeed in the project, and by fulfilling all roles six stars can be deserved (De Winter et al., 2013). Firstly, there is a 'planner' star that is reached when a planning is made, checked and, when necessary, adapted. Secondly, an 'expert' star can be deserved when pupils collect as much information as possible about something in their village by using different resources, such as the Internet, books, experts, or by visiting a company. Thirdly, to deserve the ‘researcher’ star, 5 experiments about parts of the plate have to be created. Fourthly, the ‘bookkeeper’ star can be reached when a thorough reflection is made about the materials on which money will be spend. The small budget of 100 euro requires pupils to be creative by also using recycled materials. Fifthly, a ‘builder’ star is deserved when the miniature site contains as much as possible technological applications. Finally, there is a ‘journalist’ star, for which pupils are encouraged to capture the process by making a logbook, presenting the project to parents, and so on. More information about the project can be found at the end of this introduction.

The project took a minimum of 10 and a maximum of 12 weeks (with a minimum of 20 sessions), dependent on the possibilities for the schools and teachers. The implementation of the project was supported by a conference, two workshops and two individual coaching sessions. The conference and first workshop launched the implementation of the project in the classroom. The second workshop and the two coaching sessions were organised during the implementation. In the pilot study, only the introductory conference and workshops were organised but showed to be not satisfying enough for them. Therefore, the individual coaching sessions, which took teachers’ own needs as a starting point[9], were also included in the support. During these sessions, teachers could ask specific questions about the way in which the project unfolded in their classes and about their own role therein. In order to standardise the intervention the duration and the amount of sessions were defined after the pilot study.

To answer the research questions in Studies 2 and 3, a pre- and post-measurement took place before and after the roll out of the Village@School project. To measure the quality of the interactions during S&T activities before and after the project, two standardised S&T assignments (the ‘Building a bridge’ activity and the ‘Building a tower’ activity) were developed and tested in the pilot study. During these standardised situations, a classroom observation was conducted. Early on in the pilot study it became clear that for the video-observations a camera, with a wireless microphone (attached to the teachers), was needed in order to capture teachers’ interactions with pupils. In group work settings, pupils talk with each other and so interactions with the teacher are not always clearly audible. For pupils, it was decided to make use of name stickers in order to rapidly recognise the 10 randomly selected pupils of each class and score their engagement. Next, while the pilot study demonstrated that while we could distinguish well enough between teachers regarding their teacher style (as measured with the ASOS), we chose to provide some more freedom in the teacher instructions. This was done in order to be able to appropriately determine teacher scores, as given with the CLASS. Specifically, teachers were no longer required to use an introduction and end in the lesson provided, and they could choose to use group work or not. Furthermore, in a master thesis (Smeets, 2014) concerning the analogy of both the ‘Building a bridge’ and the ‘Building a tower’ activity , it was

[9] Some teachers stated they didn’t need the support in the last coaching session anymore.

concluded that both activities were similar. This similarity was at first evaluated by an exploration of teachers' and pupils' evaluation of the difficulty level and the pleasure they had in both activities. Next, it was determined whether both situations allow an accurate measurement of the prominent variables (pupils' engagement, teacher style and class climate).

In the actual intervention research, the quality of the interactions between the teacher and pupils and among pupils was scored with the CLASS by the PhD researcher herself, and pupils' engagement (10 pupils in total) was scored with the LIS–Primary (LIS-P)[10] (Laevers, Declercq, & Jackaman, 2011) by an academic staff member of the Centre for Experiential Education (Thaline Stas). All observations were conducted live, but recorded with a video-camera. Due to time constraints, the pre- and post-observations of the second wave were not scored live with the CLASS. These observations were video-recorded by Thaline Stas and scored afterwards by the PhD researcher herself[11]. In addition, a questionnaire to measure teachers' attitudes with regard to the domain of S&T (and its teaching) was administered in the pre- and post-measurement.

Shortly after the pre-measurement, some teachers dropped out (5 out of 34 teachers). We decided to let two other 'replacing' teachers take part in the further measurements of the research. This resulted in a new sample of 31 teachers. For these two extra teachers a missing score was given in the pre-measurement. Two teachers from each school took part in the research, except for two schools: one from which only one teacher participated, and one from which three teachers were engaged (however one of these teachers dropped out). With the drop-out teachers included, 36 teachers in total participated in this research. Of all teachers, 26 were female (six of them teach in 3rd and 4th grade, one in 4th grade, six in 5th grade, two in 4th and 5th grade, ten in 6th grade and six in 5th and 6th grade) and 10 teachers were male (two of them teach in 3rd and 4th grade, one in 4th grade, four in 5th grade, two in 6th grade, one in 5th and 6th grade).

Over the course of the implementation of the Village@School project in the classroom, data were collected via different methods, but mainly via observation. The implementation of the project in the pilot study was the basis for the optimisation of the observations. Firstly, the pilot study taught us to ask teachers to work on the project with their class for a minimum of 20 lessons (of one hour each) in order to be able to spread the observations. Secondly, for the observations, they were asked to work for a minimum of one lesson on the project in order to be able to score two cycles (each 25 minutes) with the CLASS (see Study 2). Thirdly, as classes often choose to collaborate, it was asked to observe both classes with their respective

[10] While via the LIS-P a score was given on pupils' engagement, in a master thesis (Dewolf, 2016) written in function of the parallel PhD research of Veerle De Winter, also a qualitative analysis of the materials and activities on which pupils are engaged is provided.

[11] Also the observations of the pre-measurement of the first wave were again scored on the video as during the live observations in the pre-measurement of the first wave the CLASS was used for the first time in this study and some training was required.

teachers separately. Finally, the main coach of the pupils during the project had to be the teacher, and not parents, grandparents or volunteers (these were, however, allowed to help).

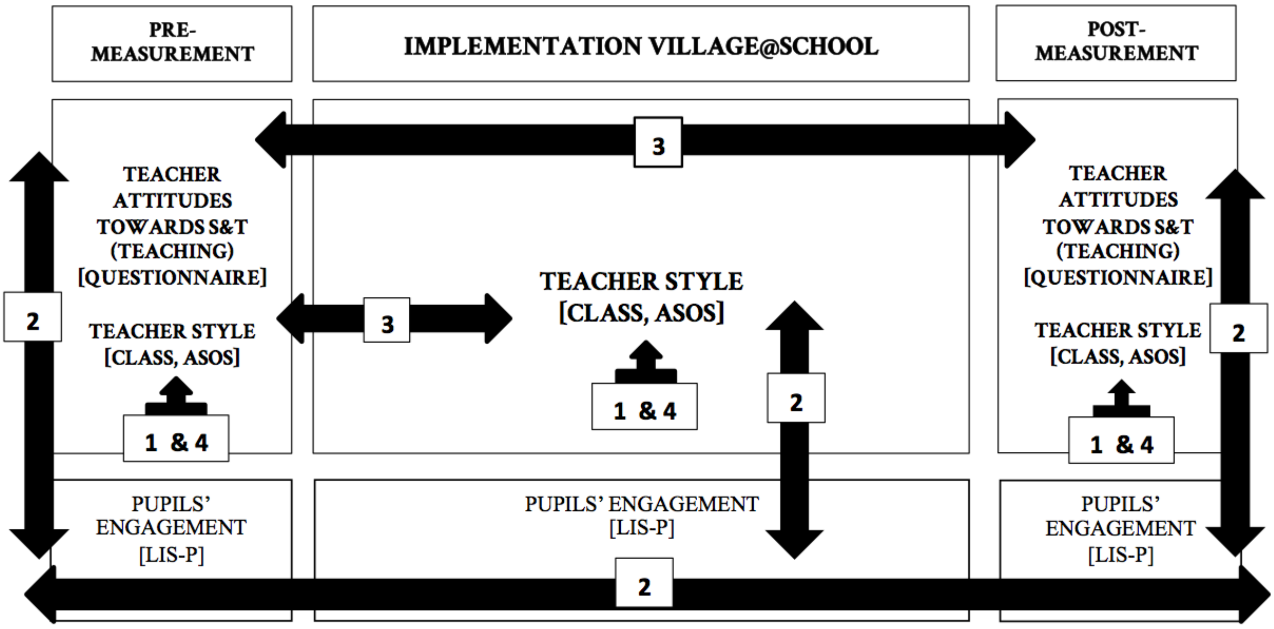
During the 3rd to 5th one-hour session of the project, and before the second workshop, a first observation (observation 1) and interview was conducted. After the second workshop, three observations (observations 2, 3 and 4) and one interview were planned (after observation 3). After observations 2 and 4, teachers were coached by another researcher of CEGO (Veerle De Winter). Observations 1 and 3 were scored live by the PhD researcher herself and another researcher of CEGO (Thaline Stas), using the same instruments as in the pre- and post- measurement. After these observations the PhD researcher conducted the teacher interviews. Observations 2 and 4 were conducted by the same colleague who had led the coaching sessions. However, these two observations had their own goals which are not relevant for this dissertation, as they were concerned with gaining insight into the development of pupils' competences in a function of the parallel PhD research. They were also video-recorded but not scored afterwards. The researchers strived to provide equal intervals between the subsequent observations for each class. Despite this, due to the absence of the teacher or the planning of other activities (such as a class trip), this was not always possible. Throughout the project, teachers kept a diary. To answer the research questions treated in Studies 2 and 3, the data collected during observations 1 and 3 was used. The teacher interview and diary data were not included in the studies of the PhD research.

It should be emphasized that this PhD research is quite unique given that an unusually large number of teachers (i.e., 31) was involved in a long-lasting intervention research for eight (the first wave) and seven (the second wave) months respectively, in which the main measurements were conducted via observations in the reality of the classroom. Overall, 126[12] observations were scored as part of the data collection for this dissertation. In other studies of open-ended S&T learning environments, most of the time only a few schools and/or teachers have been observed and analysed. The importance of such an empirical study in a larger group of teachers cannot be denied, as different researchers in the field (e.g. Krajcik et al., 1998, see manuscript 3) point to the importance of gaining insight into the effective aspects of open-ended S&T learning environments, and more specifically into the way in which teachers realise high-quality open-ended S&T learning environments. An observation tool was chosen not only to evaluate teacher style, but also to measure pupils' engagement. In most studies, data with regard to pupils' engagement are gathered by obtaining the perceptions of their own engagement (e.g. Mant et al., 2007).

[12] This number also includes the pre-measurement observations of the teachers who dropped out after the pre-measurement. The CLASS scores from the first observation were used in the analyses only for those teachers who filled in the attitude questionnaire in the pre-measurement. Several of these observations were rescored on the video.

In Figure 1 the relation between the four studies reported in this dissertation is articulated. The arrows in the scheme represent the relations between the research variables. The numbers refer to the studies wherein this relation or a variable in this relation is treated.

Figure 1. Research design



The Village@School project as an intervention

After a general description of the Village@School project and its phases, an exemplary case will be described. While in the overall research design the aims and principles of the project were discussed, in the next general description light is shed on the different phases the participating classes went through while unfolding the project. After this description, there is digged deeper into the process of one class.

General description

This description is based on CEGO's previous experiences with regard to the development and the implementation of the project, the PhD researcher's own observations during the current intervention research and the information available through teacher interviews and diaries.

During the project, classes mostly went through different phases. At first, they brainstormed about what they were going to build. While the project is called 'Village@School', the pupils were free to design any site (such as a holiday park or a theme park). One class connected this choice to the actuality: the pupils made a miniature version of 'Uplace'[13], a recently planned shopping centre in Flanders that got a lot of attention in the press because of its environmental impact. Teachers were encouraged to introduce the project to pupils in an imaginative way, for example by writing a letter from the municipality in which pupils were requested to build the site or by asking the mayor of their village to introduce the challenge (sometimes the mayor was willing to come). The latter could be kept pretty open, in order for pupils to still have enough room for initiative. Once pupils had decided what kind of site they wanted to develop, they could draw a detailed plan. Examples of constructions were a sewer system, a bridge, a big wheel, playground equipment and streetlamps. According to teachers, pupils discussed the facts and figures of the plan from the start (such as the names of the shops). Most of the time pupils had plenty of ideas, and so a selection had to be made. They often did not 'automatically' choose aspects of the site relating to S&T or did not directly connect them to S&T. Some teachers were afraid that some pupils were only 'tinkering'.

[13] Uplace is an international lifestyle/real estate group whose aim is to breathe new life into towns by establishing innovative projects." (www.uplace.eu/en/about-uplace/)

In a class discussion, groups of pupils were formed to focus on the design and building of one or more related aspects of the site. At that moment, it was often also decided which pupils were going to be the 'expert' for a particular part of the site. A planning schedule was made for a few weeks, and was followed up by one or more 'planners'. The groups of pupils gathered information for their constructions, explored how the constructions work in books or via the Internet, contacted experts and/or companies to help them with understanding certain mechanisms, explored the materials that were needed and made sketches of how the construction would look like on the standard plate (of 122 cm on 244 cm). Often in whole class discussions, the different groups presented the state-of-affairs to their peers. It was also necessary to fit the separate constructions on the plate. A particular challenge in all classes was the drawing and making of the constructions to scale. The organisation of the groups and their responsibilities was not always self-evident, as teachers experienced that they had to guide pupils in this process.

Phases could not be strictly separated from each other – as working on the project was a circular process – but mostly after a few Village@School sessions pupils no longer only 'thought' but also started 'doing'. They conducted experiments (such as how to make electricity, how to make purify water, how to make a bell, and so on), designed and started to build (for example, constructing an elevator). Experts were sometimes involved (as seen in a class trip to the Aquafin[14] company and the organisation behind the real Uplace, or an electrician coming to the class). According to teachers, some ideas, plans and experiments succeeded (for example, making pure water in an experiment with sand and shingle), which increased enthusiasm among pupils and prompted them to continue working on the project at home. In general, pupils were excited throughout the project, but some pupils also got frustrated when they did not find the necessary information or materials to build their planned construction. Throughout the process, plans were sometimes changed or particular parts were left out because pupils were stuck or no longer convinced of the viability of the construction. Teachers remarked that it was difficult for some pupils to come up with new ideas throughout the trajectory or to search for materials outside the class. Sometimes they stated to struggle with re-motivating pupils and some of them experienced that they did not have enough expertise in the fields of S&T to help. Teachers experienced help from experts in S&T or non-experts (e.g. (grand)parents), colleagues) as supportive.

At the end of the project, the miniature site was finalised and presented in a self-selected way to the other classes of the school, the directory board and the parents.

[14] Aquafin is an organisation that aims to expands, operates and pre-finances the wastewater treatment infrastructure in Flanders (www.aquafin.be)

The story of one class

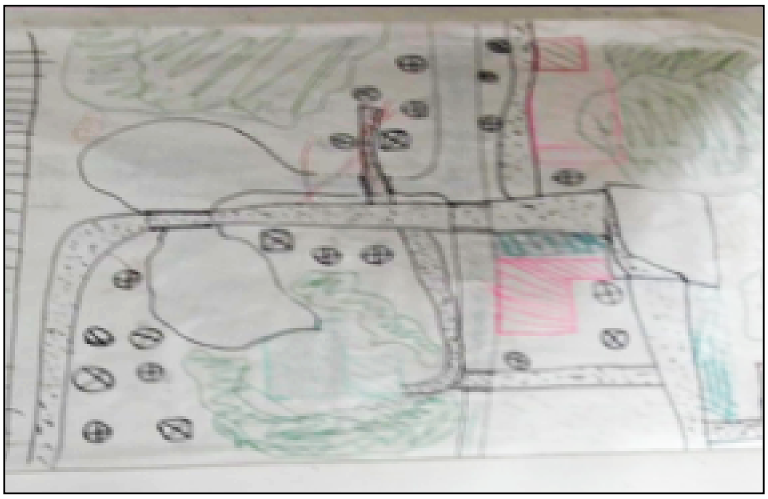
Class X, composed of 3rd and 4th grade pupils, was challenged by an architect living in the Netherlands – the brother of the headmaster – to design a new village in the neighborhood of Diest (the town in which class X's school[15] is situated) at the request of the mayor of his village. The pupils of class X were not only asked to help with the design, but had also to imagine that they would become the future citizens of that village.
At first, they visited the location, took pictures and carried out some assignments in situ. Once they were back at school, they drew a ground plan. Because they would be the future citizens of the village, all pupils chose a profession they would exercise once they lived there. One pupil wanted to be a superintendent; another wanted to start a pizzeria; still another wanted to be a hairdresser. When discussing these professions, some problems as well as opportunities arose.

A factory, for example, occupies a lot of space and they also had to think about the environment. Little by little they came to the idea of holding a small group responsible for watching over the environmental effects of particular decisions with regard to the village. When they started thinking about how their house would look like, soon the superintendent had drawn a plan at home and told the rest of the group that he 'only' needed an area of four times the classroom. He argued that the other citizens would need his factory and that he would make use of solar panels. Sharing their ideas about professions also offered the potential for opportunities. The idea to build a cultural centre, in which you can go to the theatre and also eat pizza, soon arose. As such, an actor, an artist, a restaurateur and a waitress could be employed and less space would be occupied. To support the design of their houses, two workshops – one on Gaudi given by a parent – and a visit to the city museum were organised to integrate art into the design of their houses.

The awareness grew that for some constructions, for instance the playground of the school, other decision makers than the group responsible for the environmental effects were needed. The question arose: what, in fact, do we know about decision making? In the end, elections were organised for which pupils could stand. An expert in the field – a parent who is a town councillor, – came to the classroom to talk about the mayor's responsibilities and those of the different elected local chairmen. During the elections, some pupils anxiously awaited the results. The elected mayor, chairmen and police officer were responsible for their scope and for reaching the stars. While they held meetings about the ground plan of the village and made general decisions, the other pupils – town councillors – could also contribute to the discussions.

[15] a Freinet school

Image 1: The sketched plan of the Village

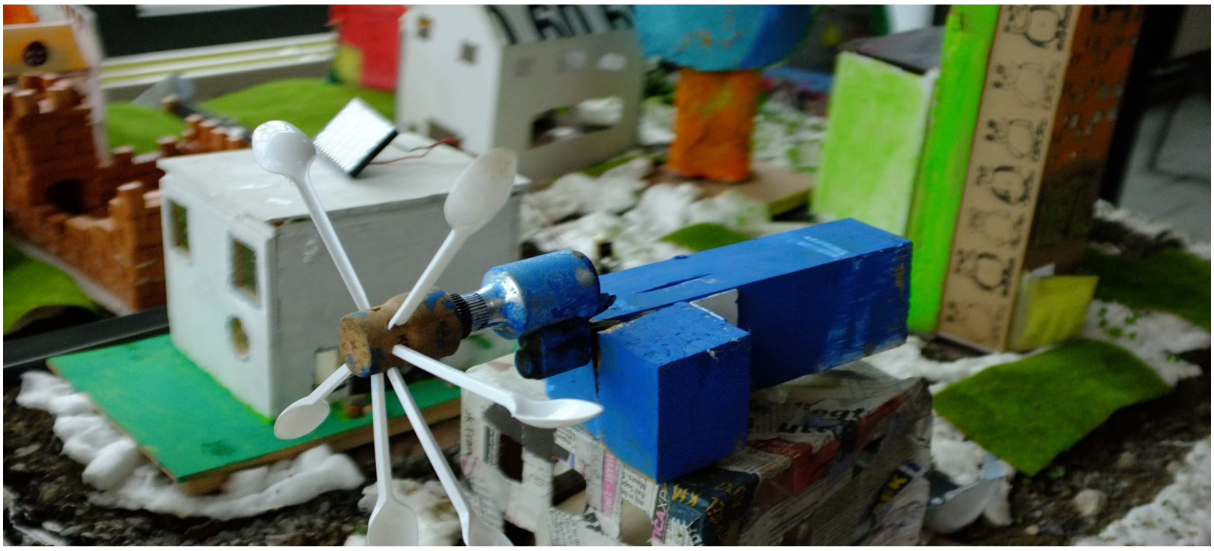


In the next step, pupils made the house they had in mind for themselves, including a front view and side-elevation, and reflected on the kind of materials to be used. Different pupils brought materials from home, such as recyclable material (polystyrene, boxes), as they wanted to build ecological houses. During this process, some questions arose. How can I make a door? What should the size of the door be so that our little man, whom we chose as the standard, can go through it? How do I fasten two things together? After the first individual coaching session for the teacher, the teacher decided to explain the working of hinge, stitching and putting elements in proportion. In the next mathematics class, they calculated how much wood they would need given the dimensions of their houses. After a long phase of brainstorming on the design of the village and the houses, they were ready to build. New problems arose, such as how the different views of the house did not fit together in practice, and for which solutions had to be found. Every Monday a parent came to the class to help pupils with their constructions.

They also reflected upon what the teacher called 'large technical problems', like wind energy, water energy and solar energy. Together with experts (the headmaster, parents, experts of the 5th and 6th grade), three groups of pupils got to work on these themes. Various windmills, a water mill, a sewer system and a solar panel were integrated into the village. As well as this, the question arose of how the different houses could be heated. As the father of one pupil makes loam stoves, the pupils went on a trip to the company where he works to figure out how you could build with loam. They also visited a factory that manufactures windows in order to learn more about heating and insulating. As the ecological aspect of the village plays an important role, a lot of greenery was foreseen in the village. Some pupils sewed seeds on wads of cotton and learnt how the plants could grow.

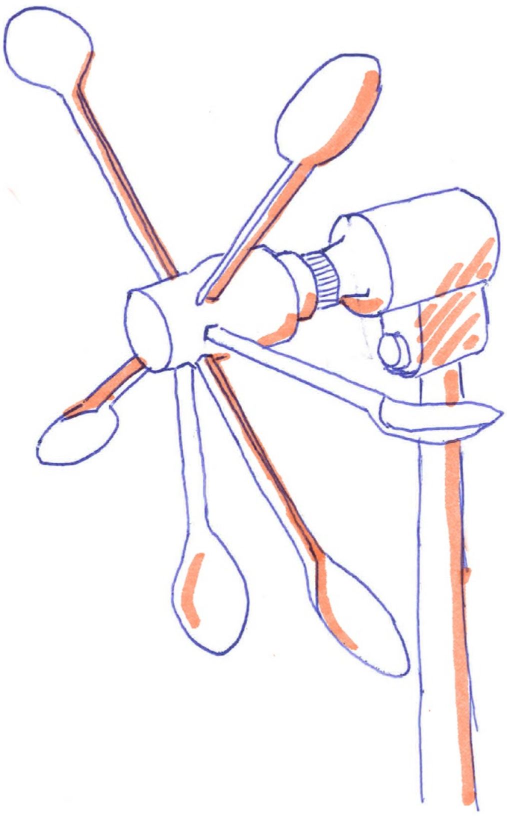
Image 2: The finished village.

Image 3: The water mill.



Finally, the constructions were given their place on the standard plate. The dimensions of the houses were drawn on the plate. A discussion with regard to the placement of the different constructions took place and the pupils noticed that the plan and the building of the scale model were not the same thing. After an observation in the post-measurement, some pupils were proudly showing off their village. Despite the teacher remarking that the experts' explanations of certain technical aspects were sometimes too difficult for some 3rd and 4th grade pupils, at the end of the project various pupils could fluently explain the working of the technical mechanisms in their village. A pupil proudly explained that an alarm in a house was set by pulling a little handle. Another pupil, who was the police officer, added that by doing so the policeman was informed when a potential criminal tried to break in. Some pupils showed how a windmill on a roof worked because of a battery in a closed circuit; using the process given in a class about the working of a battery. The sewer system, with real water, and the working of the streetlamp connected to a solar panel on the roof of a house were demonstrated. One girl clarified that the water mill works on a dynamo; they had explored the working of the dynamo on a bike in the sports hall. She explained that they were not able to let the sails turn around very quickly because a strong current of water would be needed. They tried out the water mill under a running tap, but that did not work.

It may have become clear that not only S&T was central in the implementation of Village@School in this case; it covered almost all areas of the national curriculum. As such, this case is a good example of an 'emergent curriculum' (Laevers, 2011) in which the opportunities to learn are not known beforehand, but emerge throughout the process. The teacher of this class, in comparison with teachers of other classes, proved to be exceptionally creative in the integration of other goals in the project. Moreover, a lot of different experts could contribute to the implementation of the project, and the teacher was happy to appeal to them for help. In some schools, however, the culture of helping by parents or colleagues and/or opportunities to visit companies in the neighbourhood were less present. While the pupils in this case had room for initiative, a part of the responsibility was taken by the teacher as she, by mutual agreement with the headmaster and the architect – the latter being the challenger –, decided about the different larger steps to be taken. It was, for example, decided to include a brainstorming phase long enough in duration. An architect also takes enough time for the design process. In other classes, the effective building often started earlier, still alternating with periods of brainstorming, and pupils discovered problems gradually. Finally, the teacher in this case saw her role mainly as that of an observer, who is aware of the problems facing each individual pupil in the project of and also their wishes for the design and construction of (parts of) the village. On the basis of these observations, she tried to coordinate the activities so that every pupil could find something to suit him.



Study 1: Investigating the quality of project-based science and technology learning environments in elementary school: a critical review of instruments

Published in *Studies in Science Education* 52:1 (2016): 1-27

Keywords: science and technology learning environments; project-based learning environments; primary/elementary school; instruments; data collection

Abstract

This paper provides a systematic review of instruments that have the potential to measure the quality of project-based science and technology (S&T) learning environments in elementary school. To this end, a comprehensive literature search was undertaken for the large field of S&T learning environments. We conducted a horizontal bottom-up analysis of the aspects measured by the retrieved instruments and their operationalisation. We distinguish 11 components. The most frequently evaluated components are prior knowledge and backgrounds, connection with reality, science as inquiry and level of initiative and group work. Overall, the results suggest a considerable diversity in the operationalisation of the components found. Particularly, for connection with reality, science as inquiry and level of initiative and group work, this is related to (1) the object of measurement (e.g. variety in aspects evaluated) and (2) the extent to which the used concepts are clarified. Consequently, some scales, items and questions were found to be a closer fit with aspects of project-based learning environments than others. Additionally, most of the retrieved instruments cover science and not technology or project-based education. This review can be used when searching for a scale, item or question to measure particular aspects of S&T learning environments.

Introduction

In order to cope with rapid scientific and technological developments in society, it is important that students start developing competences in the fields of science and technology (S&T) at an early age. Depending on the conceptualisation of S&T and one's view on the relationship between these two domains, different kinds of learning environments are set up and studied, among them inquiry- (Furtak, Seidel, Iverson, & Briggs, 2012), design- (Fortus et al., 2004) and project-based learning environments (Barak & Doppelt, 2000; Barak & Raz, 2000). With the concept of 'learning environment', we capture both the physical context (e.g. learning materials) as well as other qualities of the instructional approach (e.g. kind of questions teachers ask, how classroom activities are organised, etc.) that are constituted for students to learn.

This review particularly focuses on existing instruments that have the potential to investigate the quality of project-based S&T learning environments in elementary school. While this quality is quite often studied, an overview of instruments – questionnaires, surveys, interviews and observation scales – that measure the quality of these kinds of learning environments is still lacking to date.

In principle, project-based learning environments provide a context in which S&T can, but do not have to, come together (Barak & Raz, 2000). In that way, they are more 'neutral' than both inquiry- and design-based learning environments. Science, and particularly doing science as inquiry or inquiry-based teaching, refers to engaging students in scientifically oriented questions, finding evidence and coming up with explanations (National Research Council, 1996). Technology, and particularly technology as design, does not refer to the use of digital or information technology, but to the process of 'designing' technological solutions (NDET, 2006, in Hansen, 2010; Roth, 2001). Technology is the product that includes "all the knowledge that its creation entails, through which people utilise their resources on earth so that they can first survive and then move on to raise their standard of living" (Hansen, 1997, p. 112). Despite the fact that a distinction can be made between inquiry and design, they are often perceived as interrelated (Roth, 2001). Project-based learning environments in particular seem to address this interrelatedness.

Project-based learning environments are not new and have their foundations in hands-on and discovery curricula, as opposed to traditional classroom instruction (Thomas, 2000). In his review, Thomas (2000) showed that there is a diversity of defining features and no universally accepted model or theory of project-based learning environments. On the basis of his review and the frequently recurring features in the recent literature, we propose the following six main characteristics.

In project-based learning environments, learning first starts with the provision of a challenging problem (Jones, Rasmussen, & Moffitt, 1997; Liljeström et al., 2013; Luera & Otto, 2005; Marx et al., 1994; Pucel, 1992). The problem has to 'drive' students to engage in activities needed to solve the problem (Thomas, 2000), the last serving an important intellectual purpose (Blumenfeld et al., 1991). Secondly, the authenticity of the initially provided problem is characteristic of these learning environments (Donahue, Lewis, Price, & Schmidt, 1998; Hmelo-Silver, 2004; Rahm, Miller, Hartley, & Moore, 2003; Watson, 2002). Because of the confrontation with the complexity of the real world, children are challenged to disentangle physical phenomena when seeking solutions or developing products for the given challenge (Doppelt, 2009). Furthermore, the problem at stake is related to students' everyday lives (Harel & Papert, 1991; Kafai & Resnick, 1996; Rogers, Cross, Gresalfi, Trauth-Nare, & Buck, 2011; Thomas, 2000) and because of that it has some personal relevance for them (Barak, 2004; Brandt, 1998; Bransford, Brown, & Cocking, 1999; Schraw, Crippen, & Hartley, 2006). A third characteristic consists of students' opportunity to work relatively autonomously over extended periods of time (Holubova, 2008; Thomas, 2000). Fourthly, because the amount of direct guidance from the teacher is rather limited in such learning environments, students have to work collaboratively to deal with the assigned problem (Atman, Kilgore, & McKenna, 2008; Kaldi et al., 2011; Koutsides, 2001). They are involved "in the construction and co-construction of their learning and meaning making" (Hamilton, 2003, p. 37). Fifthly, although advocates of project-based learning environments make a plea for students' autonomy, most of them also simultaneously underscore the active role of the teacher. Nevertheless, there is currently no general consensus on the characteristics of such a role (Hakkarainen, 2009; Kolodner, 2001). This lack of consensus seems linked to the fact that it is unclear how much initiative should be given to students to be conducive to their learning. In this context, the 'open framework' approach is fruitful (Laevers, 2011; Schweinhart & Weikart, 1997). Initiative from both children and teachers is high in this education model. Learners and teachers realise a shared, circular process of discovery and learning (Laevers, 2011) in which the learners codetermine the path. The teacher provides a problem, the students approach it in their own way and then the teacher adjusts his/her actions to it. In these contexts, where no specific learning goals are set, the curriculum 'emerges' (Goulart & Roth, 2010). Students – guided by their own interests, goals and curiosity – are free to develop, for example, their own experiments (Kozma, Belzer, & Jaffe, 1993), while the teacher acts as a tutor, a guide and a partner in the learning process (Barth, 1972). Finally, project-based learning environments aim to foster deep understanding rather than surface understanding (Rivet & Krajcik, 2008). Throughout the years, adherents of this approach have stressed that students have to be engaged in cognitively difficult work (Blumenfeld et al., 1991) and learn 'new basic skills' (Diehl, Grobe, Lopez, & Cabral, 1999).

Because they can be integrated into project-based learning environments, inquiry- (Furtak et al., 2012) and design-based learning environments (Fortus et al., 2004) may have aspects similar to those of project-based S&T learning environments. Therefore, we will – in a first step – provide a systematic review of the instruments used in empirical research in the larger field of S&T research at the elementary school level. Secondly, those scales, items and questions of the retrieved instruments suitable for assessing aspects of project-based S&T learning environments will be selected.

This leads to the following two research questions: (1) Which aspects are measured by the existing instruments to gain insight into the quality of S&T elementary learning environments and how are these operationalised? and (2) Which scales, items and questions fit in with the aspects of project-based learning environments that were outlined above?

To our knowledge, previous work reviewing instruments in S&T education is rather limited. Liu (2010, 2012), Fraser (2012) and Wubbels and Brekelmans (2012) give an overview of instruments for research on science education, as used in elementary to high school contexts. Although these authors have familiarised researchers in the field of science education with instruments, a conceptual analysis of the particular aspects measured by those instruments is missing. Both Liu (2010) and Fraser (2012) start from a social-psychological perspective of science learning environments (Fraser 1994), which means that they measured teachers' and students' perceptions of these environments (Moos, 1979), mostly via questionnaires. More qualitative ways of data collection, via observations, interviews and logbooks, are excluded in the overviews by Liu (2010, 2012) and Fraser (2012). Although Wubbels and Brekelmans (2012) also provide a discussion of these data collection methods, these are limited to only one aspect of the learning environment, i.e. teacher-student relationships in the classroom. Finally, Liu (2010, 2012), Fraser (2012) and Wubbels and Brekelmans (2012) only include instruments used in science learning contexts, excluding those addressing the closely related field of technology education.

Method

We conducted a systematic review (Petticrew & Roberts, 2006) of the research literature by consulting the databases ERIC (CSA/ProQuest) and LIMO (the KU Leuven's electronic database, with which one can search through 255,300,000 publications, of which the bulk stems from scholarly databases and e-journal collections, free or licenced by the university and another part from the KU Leuven libraries, other scientific institutes and publications by KU Leuven researchers[16]). As we aimed to get a broad view on the data collection instruments used in the research on S&T in elementary school, the search terms were technological design; science and technology projects; science and technology projects, design or inquiry learning in combination with measurement instruments/research instruments/data collection; science teaching in combination with measurement instruments/data collection/research methodology; design education in combination with measurement instruments/data collection; science in combination with measurement instruments/research instruments/data collection/rating scale(s)/observation scale(s)/instructional effectiveness; and hands-on science in combination with research methodology (in LIMO, 'elementary/primary education' was used as a search term together with each combination; in ERIC, it was possible to select on the level of education in the advanced search tool). Among these, the search terms referring to the broader fields of design and technology, on the one hand, and to science or inquiry, on the other and finally to projects in S&T were used in combination with search terms related to the collection of data, with the word 'instrument' as the most often used search term. These terms produced a total number of 1692 hits (1365 in ERIC (CSA/ProQuest), 327 in LIMO). We reduced this sample by reading abstracts and, when necessary, full papers, eliminating studies that did not satisfy the following inclusion criteria. Firstly, the studies had to be published between January 1990 and February 2014. We could select this criterion beforehand in ERIC (CSA/Pro Quest); however, we could not do so in LIMO. Secondly, only studies with elementary/primary education as the main target group were included. Thirdly, sources were excluded if it was clear from the article that only certain programmes and/or student outcomes such as attitudes, conceptual understanding, perceptions about the content of science, etc. and/or variables influencing the teacher's approach (e.g. teachers' attitudes, beliefs, etc.) were measured, and no connection was made with aspects of the learning environment. However, it was not clear in some articles and abstracts whether instruments measured (pre-service/ in-service) teachers' beliefs or their actual teaching. When the authors of the article gave indications that the instrument could be used to evaluate the quality of the learning environment, it was included. Fourthly, articles that evaluated the quality of computer-supported learning environments, in which the teacher's role was negligible or even completely absent, were also ruled out. Because of our focus on natural class situations, we decided to include only those articles in which the teacher

[16] http://bib.kuleuven.be/english/sbib/searching-our-collections/limo

organises the learning environment. Fifthly, we excluded articles in which the data collection only consisted of observations, interviews and/or audiotaping and in which no observation scales or interview protocols nor specific questions were used or where no scales or domains were given. An example of a publication that was finally excluded because it didn't fulfil the criteria is the article by Buaraphan (2012). When checking the 'data collection' section of this study, we found that a questionnaire was used to measure teachers' conceptions of the nature of science (NOS), in addition to classroom observations, interviews and collection of related documents (e.g. lesson plans, handouts and worksheets) in order to collect data about the way in which the teacher embedded NOS in teaching about astronomy and space. The questionnaire is not useful in the light of the purpose of this review as it measures teachers' conceptions (criterion 3); the classroom observations on the contrary measure teachers' practices, but were conducted with field notes and audiotape recorder and no observation protocol/scales were used (criterion 5). Finally, the semi-structured interview also aims to evaluate what the teaching of astronomy and space looks like, but with very broad and open questions like 'what are the strengths and weaknesses in your teaching?' and 'did you embed NOS in your teaching?' (criterion 5). When the authors refer to other articles in which the instrument is used, and when it was necessary to consult them to get more information about it, or to other instruments than the one(s) they used in their article, references for these instruments were also included when they fulfilled the criteria. The application of these five selection criteria resulted in a data-set of 38 research articles, which were used for the current contribution (the articles are marked with an * in the reference list).

We first conducted a vertical or within-case analysis (Miles & Huberman, 1994) of each of the 38 research articles included in our data-set. The unit of analysis was the data collection instrument. The vertical analysis led to a classification scheme in which the data of each individual instrument concerning the following seven aspects was organised and presented: (1) the kind of instrument (observation instrument, questionnaire, interview, etc.), (2) the general concept measured in the instrument, (3) the scales/dimensions of the instrument, (4) whether the instrument is used in a science or a design educational context, (5) who used the instrument (external evaluator, teachers, etc.), (6) the school levels on which the instrument can be used (at least elementary education) and (7) the psychometric properties (validity and reliability). Aspect 3 was most essential in order to answer our two research questions; aspects (1), (2), (4), (5), (6) and (7) were chosen because they provide additional information that can help researchers in choosing particular instruments for the data collection in their studies. For example, researchers who conduct studies on other educational levels than elementary education can profit from this review since instruments that can be used at these levels were also discussed. When, for instance, only searching for a survey in order to measure the quality of the S&T learning environment, one can see at a glance which instruments are surveys, and thus interesting to consider.

The two research questions above were answered by conducting a horizontal or cross-case analysis (Miles & Huberman, 1994). To answer research question 1, concerning the aspects measured in the instruments and their operationalisation, we did a conceptual analysis of the scales, items and/or questions of each instrument. We looked for the different aspects of the learning environment measured in the data-gathering instruments and compared the operationalisation of these aspects in each instrument. Therefore, a bottom-up analysis was conducted, which means that the starting point for making categories of components – as aspects of the learning environment – was formed by the mere descriptions of the instruments, their scales, items and questions. Since every categorisation inevitably involves a reduction of reality and the instruments, scales, items and questions may simultaneously refer to another component, too, we decided to place them under a specific component category based on their main connection to that component (this analysis was conducted by the first author). In case of doubt, we consulted each other. Regarding question 2, concerning the instruments, scales, items and questions that fit in with the aspects of project-based learning environments, we drew a conclusion about the compatibility of the instruments, scales, items and questions with the six aspects of project-based learning environments as described in our theoretical framework (i.e. challenging problem, authenticity of the problem, level of initiative, working collaboratively, active role of the teacher and deep understanding). The results of this horizontal analysis aimed at answering our two research questions are presented below. The results section starts with some information concerning some general aspects of the retrieved instruments.

Results

General information about the instruments

Information concerning general aspects of the retrieved instruments is provided in Table 1. The table shows that the bulk of the instruments (17) consists of a survey or questionnaire, another 11 instruments are observation protocols or guidelines, two instruments are used for document analysis (lesson plans) and only four instruments are interviews. Two self-reflection instruments and one coding scheme were also retrieved. These instruments were almost exclusively applied in science educational research – a few are also used in the context of mathematics – but none of them was made for the field of technology education. Six instruments – mainly questionnaires and surveys – can be filled in by school students and 13 by teachers. The other instruments are meant for use by external evaluators (observers) and interviewers. All instruments are made for or used on the elementary school level, but quite a lot of instruments are also suitable for higher educational levels. Information concerning the psychometric qualities of each instrument was not always given. More information was available with regard to reliability than with regard to validity.

This table makes it possible to get a quick view on suitable instruments, scales, items and questions with regard to one's own research aims. After having made a first selection, it can be useful to get more insight into the aspects measured. For this, we refer to Table 2. The numbers of the instruments in Table 1 correspond to the numbers used in Table 2.

Variables addressed by the instruments and their operationalisation

We looked for similarities and differences between scholars' views on the characteristics of powerful learning environments for S&T learning environments. Table 2 shows which components were evaluated. Using a bottom-up approach, we identified the following 11 components: anticipating students' prior knowledge and backgrounds, the connection with reality, science as inquiry, interesting activities, group work, level of initiative, teachers' interventions, teachers' content knowledge, use of new technology, fostering understanding and evaluating students' understanding. An instrument is situated in one or more categories either based on the conceptualisation of the instrument as a whole or on the conceptualisation of one or more specific scales, items and/or questions.

Table 2 allows the aspects measured by the instruments to be viewed at a glance. This is indicated in the columns referring to the components. The table is intended to be a point of departure to examine the specific operationalisation of the instrument in the component subsections of this article and using that to determine the extent to which this fits with one's research aims (starting on p. 37). The numbers corresponding to the instruments make it possible to swiftly recognise the instruments throughout the manuscript.

Before starting the overview of the different components, we want to note that there are some general and some specific scales that are difficult to classify in one of the 12 categories. The most important ones are described below.

Firstly, two instruments [26 and 31] contain a scale or dimension referring to 'instruction' in general, but no detailed information about them was available. In a similar way, one instrument [31] measures whether 'the resources available in this lesson contributed to accomplishing the purposes of the instruction' and examines whether the 'design for future instruction takes into account what transpired in the lesson'. Secondly, only one instrument [16] – with the Lesson Segments scale – more generally evaluates the quality of different lesson stages (e.g. introduction). Just one instrument [37] evaluates how often teachers assign homework. Thirdly, two instruments have an item that is very specific as it assesses the effects of a programme teachers attended [28 and 29]. Similarly, one instrument evaluates the frequency with which particular materials are used during the implementation of a programme [37]; another similar one contains an implementation dimension that determines whether the instruction was consistent with the underlying approach of these materials [31].

Table 1. General Information on the Retrieved Instruments

Instrument	Considered articles in which the instrument is used	Kind of instrument	What's measured in general?	Scales/dimensions	Science and/or design (technology) education research	Instrument used by	School level	Validity	Reliability
1. The 5E Lesson Plan Scoring Instrument/5E Lesson Plan (5E ILPv2) rubric (Goldston et al., 2010)	Goldston, Dantzler, Day and Webb (2013)	Document analysis (lesson plans)	Pre-service teachers' abilities to create written inquiry-based 5E lesson plans	5 scales (engage, explore, explain, elaborate, evaluate)	Science and design education	External evaluator	Elementary education	Construct validity for five factors (85.5% of the variability explained) Content validity	α=.94 (engage scale), α=.99 (explore scale), α=.96 (explain scale), α=.97 (elaborate scale), α=.95 (evaluate scale); α=.95(total scale)
2. Survey (Forbes & Davis, 2008)	Forbes and Davis (2008), Forbes and Davis (2011)	Survey	Pre-service teachers' development of curricular role identity for science teaching through their use of curriculum materials	4 dimensions (curricular role identity for general use of science curriculum materials, for scientific inquiry, for curriculum materials use in context and for teacher learning from curriculum materials)	Science and mathematics education	Pre-service elementary teachers	Elementary education		α = 0.79 within dimension 1 α = 0.89 within dimension 2 α = 0.71 for dimensions 3 and 4 (Forbes & Davis, 2008)
3. Constructivist Learning Environment Scale (CLES) (Taylor & Fraser, 1991)	Taylor and Fraser (1991), Roth and Bowen (1995)	Survey	The extent to which students perceive their learning environment as consistent with a constructivist epistemology	4 scales (autonomy scale, prior knowledge scale, negotiation scale, student-centeredness-scale)	Science education	Elementary school students	Elementary and secondary education	Satisfactory discriminant and predictive validity in secondary schools (Taylor & Fraser, 1995)	α=.52 -.72 (Roth & Bowen, 1995) satisfactory internal consistency in secondary schools (Taylor & Fraser, 1995)

Instrument	Considered articles in which the instrument is used	Kind of instrument	What's measured in general?	Scales/dimensions	Science and/or design (technology) education research	Instrument used by	School level	Validity	Reliability
4. Constructivist Learning Environment Survey (CLES) (Taylor, Dawson, & Fraser, 1995; Taylor, Fraser & Fisher, 1993; Taylor et al., 1993, 1997), the CLES 1(30) (Taylor et al., 1995, 1997), the CLES2 (20) (Johnson & McClure, 2004)	Johnson and McClure (2004); Türkmen (2009)	Survey	Science classroom learning environments (as organised by pre-service and in-service teachers)	5 scales (personal relevance, uncertainty, critical voice, shared control, student negotiation) (Taylor et al., 1997)	Science education	Elementary, middle and high school (pre-service and in-service teachers and students)	Elementary, secondary and high school education	Construct validity for five scales (Taylor et al., 1997)	α=.88 for the overall CLES 1(30); α=.93 and α =.94 for the overall CLES 2 (20)
5. Questionnaire (Tao, Oliver, & Venville, 2013), nine items were selected from the TIMSS 2007 student questionnaires (http://timssandpirls.bc.edu/TIMSS2007/context.html)	Tao et al. (2013)	Survey	The experiential curriculum, or the experiences during science instruction as perceived by students	No specific scales, students evaluate how often they participate in science activities (e.g. designing a science experiment or investigation)	Science education	Elementary school students	Elementary education		
6. Curriculum Evaluation Tool, created by AAAS Project 2061 (American Association for the Advancement of Science, 2002) and adapted by Wilson	Baxter et al. (2004)	Document analysis (lesson plans)	If curriculum items of pre-service teachers' lesson plans meet the standards of each AAAS-cluster	5 clusters: providing a sense of purpose for student and the teacher, taking account of student ideas to inform instruction, engaging students with phenomena, developing and using scientific ideas and skills during instruction and examining assessment of progress	Science education	External evaluator	Elementary and secondary education		
7. Changes to Teaching Practices Questionnaire	Pop, Dixon, and Grove (2010)	Questionnaire	Changes to teaching practices of in-service teachers after attending the Research Experiences for Teachers (RET) program	No scales, but items (e.g. 'I made some general changes to the instructional strategies and my teaching style', 'my class is more student-centered', 'I do more experiments than I did before', 'I do more hands-on activities')	Science education	In-service elementary, middle and secondary teachers	Elementary, middle and secondary education		
8. Semi-structured interview	Pop et al. (2010)	Interview protocol	Changes to teaching practices after attending the RET program	Some general questions (including questions about professional development), some questions about describing changes in teaching practices after attending the RET program, a set of questions about classroom science instruction (e.g. types of instructional science activities)	Science education	Interviewer	Elementary, middle and secondary education		Initial interrater reliability between two coders: 90%

9. Classroom Observation Guideline, based on other instruments (Newmann, Secada, & Wehlage, 1995; Horizon Research, 2000; Piburn & Sawada, 2000)	Luykx and Lee (2007)	Observation guideline	Different aspects of classroom practice	8 scales (scientific understanding, scientific inquiry, scientific discourse, teachers' knowledge of science content, diversity of cultural experiences and materials, students' home language, scientific authority, linguistic scaffolding to enhance meaning, which are grouped in three categories (constructs of science learning, constructs based on students' linguistic and cultural knowledge and constructs that bridge the two domains).	Science education	External evaluator (observer)	Elementary education	Difficult to check because of scarcity of similar instruments whose validity has been well-established	r=.60 -.84 (initial); r=. 81 (later)
10. Classroom observation guideline (adapted from the observation scales for authentic instruction (Newmann & Associates, 1996))	Lee, Hart, Cuevas, and Enders (2004)	Observation guideline	Teaching practices	Four scales: scientific understanding, scientific inquiry, scientific discourse, teacher's knowledge of science content	Science education	External evaluator (observer)	Elementary education	More valid when more observations can be done	More reliable when more observations can be done r=.60 -.84 (initial); r=.81 (later)
11. Classroom observation guideline	Lee, Luykx et al. (2007)	Observation guideline	In-service teacher's behaviour	Three scales: students' home language, diversity of cultural experiences and materials, and culturally congruent communication and interaction	Science education	External evaluator (observer)	Elementary education		r=.60 -.84 (initial); r=.81 (later)
12. Interview	Lee, Luykx et al. (2007)	Interview protocol	Determines how in-service teachers incorporate students' home language and culture into science instruction	A set of questions regarding teachers' conceptions about the level of students' prior knowledge in science, leading to teachers' conceptions of how students' culture, home language, and SES influence science learning	Science education	External evaluator	Elementary education		
13. Questionnaire	Lee, Luykx et al. (2007)	Questionnaire	Determines how teachers incorporate students' home language and culture into science instruction	2 sets of items: one set related to the construct of students' home language and one set related to the construct of students' home culture (determination of teachers' knowledge and importance with regard to these constructs)	Science education	In-service elementary school teachers	Elementary education		α =.85 -.97
14. NSF-CETP Student Teacher Videotaped Lessons Scoring Protocol	Online Evaluation Resource Library, 2004, as cited in Levy, Pasquale and Marco (2008)	Observation instrument	Observing and assessing instruction of in-service teachers	Three topics covered: instructional practices, project evaluation, student/teacher interaction	Math and science education	External evaluator: observer	K-12		
15. Semi-structured interview	Eick (2011)	Interview protocol	Evaluating in-service teachers' science and nature-study curriculum and development and use of the outdoor classroom	8 questions (e.g. how does the environment integrate into your teaching of science?)	Science education	Interviewer	Elementary education		

Instrument	Considered articles in which the instrument is used	Kind of instrument	What's measured in general?	Scales/dimensions	Science and/or design (technology) education research	Instrument used by	School level	Validity	Reliability
16. Innovation Configurations Checklist (Hall & Hord, 1987, Chapter 5), derived from observation procedures used by Hollon, Anderson and Smith (1980)	Neale, Smith and Johnson (1990)	Observation instrument	Rating lesson features of conceptual change teaching	Three levels of implementation specified for each of 31 features of conceptual change teaching. Those 31 features were grouped in 31 categories.	Science education	External evaluator (observer)	Elementary education		α=.94-1.00
17. The Draw-A-Science-Teacher-Test Checklist (DASTT-C), derived from Goodenough's Draw-A Man Test (1926) and Chambers (1983)	Yilmaz, Turkmen, Pedersen and Huyuguzel Cavas (2007); Koch and Appleton (2007); Thomas, Pedersen and Finson (2001)	Questionnaire	Elementary pre-service teachers' perceptions of teaching science, their teaching approaches are assessed	Pre-service elementary teachers have to 'Draw a picture of themselves as a science teacher at work' and to explain this (questions: 'What is the teacher doing?'; 'What are the students doing?'). Teachers are categorised in: student-centered teaching style, teacher-centered teaching style, neither student-centered nor teacher-centered teaching style.	Science education	Pre-service elementary teachers	Elementary education	Content validity was checked via review of drawings (Yilmaz et al., 2007) Content validity ok (Thomas et al., 2001)	KR-20 = 0.82 (Thomas et al., 2001), KR-20=0.71 (Yilmaz et al., 2007) α =.82 (Thomas et al., 2001)
18. Reformed Teaching Observation Protocol (RTOP) (Piburn & Sawada, 2000; Sawada et al., 2000)	Lakshmanan, Heath, Perlmutter and Elder (2011), Piburn and Sawada (2000)	Observation instrument	Reformed teaching as teaching incorporating "constructivist, inquiry-based methods" that are grounded in mathematics and science education research and the national standards (MacIsaac & Falconer, 2002)	5 sections: lesson design and implementation, propositional content knowledge, procedural content knowledge, classroom culture (communicative interactions), and classroom culture (student-teacher relationships)	Science and mathematics education	External evaluator (observer)	K-20 education	Highly valid (Piburn & Sawada, 2000)	highly reliable (Piburn & Sawada, 2000)

19. The Science Teacher Inquiry Rubric (STIR), based on the Web-Based Inquiry for Learning Science [WBI] Instrument (Bodzin & Cates, 2002)	Bodzin and Beerer (2003)	Self-reflection instrument Observation instrument	Teacher behaviors, against the light of the five features of classroom inquiry and their variations based on the amount of learner self-direction and direction from materials (NRC, 2000)	5 categories: learners are engaged by scientifically oriented questions; learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions; learners formulate explanations and conclusions from evidence to address scientifically oriented questions; learners evaluate the explanations in light of alternative explanations, particularly those reflecting scientific understanding; learners communicate and justify their proposed explanations	Science education	Elementary school teachers External evaluator (observer)	Elementary, secondary and high school education	Content validity provided	r=.58 (self-assessment) r=1 (observation)
20. SPRinG-Teaching Observation Protocol (S-TOP)	Howe et al. (2007)	Self-reflection instrument	Ratings of the group activity (group structure, teacher input and student interaction) that took place across full teaching sessions	31 scales: four 'learning context' scales, seven 'activities and tasks' scales, nine 'role of adults' scales, eleven 'group interaction' scales	Science education	External evaluator (observer)	Elementary education		
21. My Class Inventory (MCI), adapted from the Learning Environment Inventory (LEI) (Fraser & O'Brien, 1985)	Houston, Fraser and Ledbetter (2008)	Questionnaire	The effectiveness of instruction, the classroom environment	Five scales (classroom environment dimensions): cohesiveness, friction, difficulty, satisfaction and competition)	Science education	Elementary school students	Elementary education	Validity confirmed (Mink & Fraser, 2005; Sink & Spencer, 2005; Goh & Fraser, 1998; Majeed, Fraser, & Aldridge, 2002; Fraser & O'Brien, 1985) Discriminant validity for the friction, competition and cohesiveness scale:.30 -.32 (pretest),.25-.38 (posttest 1), and .36 -.38 (posttest 2). Strong factorial validity for a revised four factor version of the MCI (Houston et al., 2008; Sink & Spencer, 2005)	α = 0.61-.76 (pretest), α=.53 -.78 (posttest 1), α = 0.69 -.80 (posttest 2)
22. Survey questionnaire	Braund and Leigh (2013)	Survey	Changes in teachers' classroom actions that might have been due to professional development on using talk-related science activities (8 months of project intervention in	Reporting of use of group work, organisation of groups in science lessons, the most important outcomes of group work in science, an example of the most successful experiences of group work in science lessons and what might help or hinder students' progress in engaging with group work. In the second application of the questionnaire also questions about changes in group work are posed.	Science education	Elementary school teachers	Elementary education		

Instrument	Considered articles in which the instrument is used	Kind of instrument	What's measured in general?	Scales/dimensions	Science and/or design (technology) education research	Instrument used by	School level	Validity	Reliability
23. Survey	Rennie, Goodrum and Hackling (2001)	Survey	Students' responses to, and perceptions and ideas about science at school	Demographic data, items with the general heading 'how often do these things happen in your science lessons?'; items with the general heading 'How often are these things true for your science lessons?'; four or three open-ended questions (secondary respectively primary questionnaires)	Science education	Elementary and secondary school students	Elementary and secondary education		
24. Questionnaire	Braund and Leigh (2013)	Questionnaire	Recording data reported by students on three parameters: frequency of learning experienced, students' perceived self-efficacy of these experiences and their attitudes to school science	Three parameters: frequency of learning experienced, students' perceived self-efficacy of these experiences and their attitudes to school science	Science education	Elementary school students	Elementary education		
25. Time-sampling methodology	Howe et al. (2007)	Observation instrument	Ratings observations of the group activity *in situ*	12 behavioral categories to score during the sampled periods: four categories relate to the social context in which students were located; 8 categories cover key aspects of dialogue	Science education	External evaluator (observer)	Elementary education		Inter-researcher agreement over the categories ranged from 82% to 100% (M= 92%)
26. New Constructivist Learning Environment Scales	Wang and Lin (2009)	Questionnaire	Identifying various dimensions of the learning environment	Four dimensions: attitude concerning instruction, interaction, scientific inquiry, and understanding	Science education	Elementary and middle school students	Elementary and middle school	Satisfactory	Satisfactory
27. Survey	Marshall, Horton, Igo and Switzer (2007)	Survey	Measuring beliefs about and use of inquiry in the classroom	16 demographic questions, 17 items measuring beliefs about inquiry instruction, content standards and support structures and ten items measuring how often teachers engage in inquiry and frequency that career connections are made with the curriculum	Science and mathematics education	Elementary, middle and high school teachers	Elementary, middle and high school education	Validity for self-efficacy scale ok (varimax rotation)	α=.87

28. Survey 1 (pre and post)	Banks, Elser and Saltz (2005)	Survey	Recording a change in teachers' knowledge acquisition, in the way they deliver their science curriculum and determining if they accomplish what they thought they would after attending the Ecology Explorers	Two categories of items: teacher knowledge and communication; student research	Science education	Elementary and secondary school teachers	Elementary and secondary education		α =.82 (post survey)
29. Survey 2 (post)	Banks et al. (2005)	Survey	Determining the factors that appear to facilitate or impede the implementation of protocols into curriculum and the extent in which the EE program methodologies are used to enhance teaching practices in inquiry, research, student experimentation and the integration of mathematics	18 items centering on frequency of use of the EE program methodologies, support and interest, use of the EE program methodologies in teaching inquiry, and teaching practices	Science and mathematics education	Elementary and secondary school teachers	Elementary and secondary education		α =.91
30. Questionnaire	Lee, Lewis, Adamson, Maerten-Rivera and Secada (2007)	Survey	Teachers' knowledge and practices in teaching science to English Language Learning students	4 scales: teacher knowledge of science content scale, practice in scientific understanding scale, practice in scientific inquiry scale, practice in English language development scale	Science education	Elementary school teachers	Elementary education		α =.72-.90
31. 2003-2004 and 2005-2006 Local Systemic Change Classroom Observation Protocol (Horizon Research Inc., 2003)	Levy et al. (2008), Horizon Research (2003), Shymansky, Wang, Annetta, Yore and Everett (2010)	Observation instrument	Record and rate mathematics and science lessons from classrooms in Local Systemic Change districts (designed to capture all of the aspects of classroom instruction)	Four categories: design, implementation, mathematics/science content, classroom culture	Mathematics and science education	External evaluator (observer)	Grades 6-12 and grades K-8		

Instrument	Considered articles in which the instrument is used	Kind of instrument	What's measured in general?	Scales/dimensions	Science and/or design (technology) education research	Instrument used by	School level	Validity	Reliability
32. Coding scheme based on Shulman's (1987) stages of pedagogical reasoning	James and Sharmann (2006)	Coding scheme	Determines how instruction in the use of analogies might influence the teaching performance of pre-service teachers	Six stages (Shulman, 1987): comprehension, transformation, instruction, evaluation, reflection and new comprehension	Science education	External evaluator	Elementary education		r = 0.94
34. Postobservation interview	Lee, Lewis et al. (2007)	Interview	Teachers' reflections on their practices during the observed lessons	The interview protocol contains four opening questions on teaching practices to promote scientific understanding and inquiry and on teaching practices to support English language development. These are followed by probes using key examples from the day's lesson.	Science education	Interviewer	Elementary education		Interrater agreement: 90%
35. Survey	Newman et al. (2012)	Survey	Possible change in classroom practice after professional development	Questions in 11 domains: professional development, instructional time, student assessment, technology, teacher background, equipment and materials, instructional strategies, planning, collaboration and support, student engagement and teacher content knowledge and implementation	Science and mathematics education	Elementary and secondary in-service teachers	Elementary and secondary education		
36. CETCP Core Evaluation Classroom Observation (Lawrenz, Huffman, & Appeldorn, 2002)	Levy et al. (2008)	Observation instrument	Classroom practices	Type of instruction, students' engagement and cognitive activity of students (time-sampled observational method)	Science education	External evaluator (observer)	Elementary education		
37. Local Systemic Change Teacher Questionnaire	Young and Lee (2005), Horizon Research (2003)	Questionnaire		Demographic data (gender, race/ethnicity, science coursework, years of teaching experience), information regarding the setting in which the teachers teach science (grade level, class size, duration of typical science lesson, number and types of units taught, and instructional approaches)	Science education	In-service elementary teachers	Elementary education		

Note: From left to right: the instrument (in the manuscript is referred to its number), the articles in which it was retrieved, the kind of instrument (survey, questionnaire, observation instrument and interview), what it measures in general, the scales, items and questions that are used, whether the instrument can be used in science and/or technology and/or mathematics education, who completes in the instrument (in-service/pre-service teachers; external evaluators (observers); elementary/secondary/middle school teachers or students); the school level contexts in which the instrument can be used (elementary/secondary/middle school) and the validity and reliability of the instrument (in case this information was available).

Table 2. Retrieved instruments: bottom-up analysis

Instrument	Author	Prior Knowledge & Backgrounds	Connection with Reality	Science as Inquiry	Interesting Activities	Group Work	Level of Initiative	Fostering Under-standing	Use of New Technology	Teachers' Interventions	Teachers' Content Knowledge	Evaluating Students' Understanding
1. The 5E Lesson Plan Scoring Instrument	Goldston, Dantzler, Day & Webb (2012)	x	x			x	x		x	x		x
2. Survey	Forbes and Davis (2008), Forbes and Davis (2011)	x								x		
3. Constructivist Learning Environment Scale (CLES)	Taylor and Fraser (1991), Roth and Bowen (1995)	x	x	x		x	x					
4. Constructivist Learning Environment Survey	Johnson & McClure (2004), Türkmen (2009)	x	x				x					
5. Questionnaire	Tao, Oliver & Venville (2013)			x		x	x	x		x		
6. Curriculum Evaluation Tool	Baxter, Jenkins, Southerland and Wilson (2004)		x	x	x		x					x
7. Changes to Teaching Practices Questionnaire	Pop, Dixon and Grove (2010)			x			x			x		
8. Semi-structured Interview	Pop, Dixon and Grove (2010)			x								
9. Classroom Observation Guideline	Luykx and Lee (2007)	x	x	x		x		x			x	
10. Classroom Observation Guideline	Lee, Hart, Cuevas and Enders (2004)			x		x		x			x	
11. Classroom Observation Guideline	Lee, Luykx, Buxton, & Shaver (2007a)	x									x	
12. Interview	Lee, Luykx, Buxton, & Shaver (2007a)	x		x				x				
13. Questionnaire	Lee, Luykx, Buxton, & Shaver (2007a)	x						x				
14. NSF-CETP Student Teacher Videotaped Lessons Scoring Protocol	Online Evaluation Resource Library, 2004, in Levy et al., 2008		x	x		x			x			
15. Semi-structured Interview	Eick (2011)		x			x						

	Instrument	Author	Prior Knowledge & Backgrounds	Connection with Reality	Science as Inquiry	Interesting Activities	Group Work	Level of Initiative	Fostering Under-standing	Use of New Technology	Teachers' Interventions	Teachers' Content Knowledge	Evaluating Students' Understanding
16.	Innovation Configurations Checklist	Neale, Smith and Johnson (1990)		x			x	x	x		x		
17.	The Draw-A-Science-Teacher-Test Checklist (DASTT-C)	Yilmaz, Turkmen, Pedersen, & Huyuguzel Cavas, 2007; Koch & Appleton, 2007; Thomas et al., 2001 (DASTT)						x			x		x
18.	Reformed Teaching Observation Protocol (RTOP)	Lakshmanan, Heath, Perlmutter & Elder (2011)					x	x	x		x		
19.	The Science Teacher Inquiry Rubric (STIR)	Bodzin and Beerer (2003)						x					
20.	SPRinG-Teaching Observation Protocol (S-TOP)	Howe, Tolmie, Thurston, Topping, Christie, Livingston, Jessiman & Donaldson (2007)					x						
21.	My Class Inventory (MCI)	Houston et al. (2008)	x			x	x						
22.	Survey Questionnaire (teachers)	Braund and Leigh (2013)					x						
23.	Survey	Rennie et al. (2001)											
24.	Student Questionnaire	Braund and Leigh (2013)					x						
25.	Time-sampling Methodology	Howe et al. (2007)					x						
26.	New Constructivist Learning Environment Scales	Wang and Lin (2009)			x				x		x		
27.	Survey	Marshall, Horton, Igo and Switzer (2007)			x								
28.	Survey 1 (pre en post)	Banks, Elser and Saltz (2005)											
29.	Survey 2 (post)	Banks, Elser and Saltz (2005)			x								
30.	Questionnaire	Lee, Lewis, Adamson, Maerten-Rivera and Secada (2007b)	x		x				x			x	
31.	2003-2004 en 2005-2006 Local Systemic Change Classroom Observation Protocol	Levy, Pasquale and Marco (2008), Horizon Research (2003), Shymansky et al. (2010)	x	x	x	x	x	x	x		x	x	x

Instrument	Author	Prior Knowledge & Backgrounds	Connection with Reality	Science as Inquiry	Interesting Activities	Group Work	Level of Initiative	Fostering Under-standing	Use of New Technology	Teachers' Interventions	Teachers' Content Knowledge	Evaluating Students' Understanding
32. Observation Framework	James and Sharmann (2006)											
33. Classroom Observation Scales	Lee, Lewis, Adamson, Maerten-Rivera and Secada (2007)	x		x							x	
34. Postobservation Interview	Lee, Lewis, Adamson, Maerten-Rivera and Secada (2007)	x						x				
35. Survey	Newman et al. (2012)	x		x		x			x		x	x
36. CETP Core Evaluation Classroom Observation	Levy, Pasquale and Marco (2008)	x										
37. Local Systemic Change Teacher Questionnaire	Young and Lee (2005); Horizon Research (2003)	x	x	x	x	x	x	x	x	x		x

Note: From left to right: the instruments, the articles in which they were retrieved and the components in which they can be situated (Prior Knowledge & Backgrounds, Connection with Reality, Science as Inquiry, Interesting Activities, Group Work, Level of Initiative, Fostering Understanding, Use of New Technology, Teachers' Interventions, Teachers' Content Knowledge and Evaluating Students' Understanding) - on the basis of their scales, items and questions. Each instrument has a number to which is referred throughout the manuscript.

Anticipating students' prior knowledge and backgrounds

This first component deals with the evaluation of teachers' anticipation of the baggage students bring with them when they come to the classroom: their prior knowledge, their interests, as well as their cultural background and educational needs.

Firstly, five instruments [1, 2, 31, 35 and 36] evaluate teachers' anticipation of students' prior knowledge. Two of them [1 and 2] view this as a way to engage students. In one instrument aimed at evaluating lesson plans [1], for example, the teacher's strategies to ascertain students' prior understanding of science concepts are integrated in the *engage* scale, together with the encouragement of students' questions, the generation of students' interest in the activities to be taught and in what follows. Three other instruments [31, 35 and 36] also address students' engagement, but it is unclear whether they deal with the promotion of students' engagement by making use of students' prior knowledge.

While not explicitly connecting students' prior knowledge to students' engagement, one survey [3] explicitly links students' prior knowledge to real life. Roth and Bowen (1995) view the item 'In this class, I think about interesting real life problems' as characteristic for the *prior knowledge* scale, which measures students' perceptions of the opportunities for meaningful integration of knowledge in the classroom.

Secondly, one instrument measures the extent to which school science is linked to students' everyday lives, but without making reference to students' prior knowledge. This instrument [4] uses the *personal relevance* scale to measure students' perception of the extent to which school science/mathematics is relevant to their everyday in- and out-of-school experiences.

Thirdly, several instruments refer to the connection of the lesson with students' cultural background. In four instruments [9, 11, 12 and 13], the *diversity of cultural experiences and materials* scale evaluates the extent to which students' cultural experiences and materials are integrated into science instruction. In another instrument [31], this aspect is evaluated with the item 'the instructional strategies and activities reflected attention to issues of access, equity, and diversity for students (e.g. cooperative learning[17], language-appropriate strategies/materials)'. In the *classroom culture* dimension of the same observation instrument, the observer can write down examples of the extent to which there is an appreciation of diversity among students (e.g. their gender, race/ethnicity and/or cultural background). Lee, Luykx,

[17] This item could also be classified in the group work component, although in does not fit in one of the categories we discerned.

Buxton, and Shaver (2007) add another similar scale in their instruments [11, 12 and 13] that assesses the extent to which the teacher communicates and interacts with students in culturally congruent ways.

Fourthly, three instruments [30, 33 and 34] – the first with the practice in *English language development* scale, the second with the *support of students' English language development* scale and the third with two questions (e.g. 'I'd like to know about the strategies that you use to promote students' English language development' and 'Do you have ESOL[18] students in your class? I'd like to know about the strategies that you use to promote ESOL students' English language development') – assess the practices to support English language development during science lessons. By cultural background, we also refer to students' linguistic competence in English and in their home language. Different studies investigate whether science activities can foster the development of the English language of English language learners (ELLs) – mostly immigrant students in the USA – and which science activities and support from the teacher are important to communicate about science (Amaral, Garrison, & Klentschy, 2002; Lee, Deaktor, Hart, Cuevas, & Enders, 2005; Rosebery, Warren, & Conant, 1992). Because this support also encompasses that teachers take into account ELL students' oral and written proficiencies in their home language (Lee, Luykx et al., 2007), instruments that measure this support are also included in the discussion of this component.

Fifthly, some instruments evaluate to what extent teachers anticipate students' other educational needs. One instrument [31] extensively assesses whether science instruction aligns with students' experience, learning styles and developmental levels. The questionnaire of the same name [37] only determines how often teachers 'allow students to work at their own pace'. The *difficulty* scale of another instrument [21] evaluates the extent to which students experience difficulty with class work.

Connection with reality

The literature shows that teachers often bring reality into the classroom when they organise S&T activities for their students. While we have already discussed an instrument that evaluates the connection of science in the classroom to students' everyday lives, some other instruments evaluate the connection with reality as such, without referring to the idea that this reality should stem from students' everyday lives. Among the 17 instruments that measure this connection with reality, 3 different ways could be identified. Remarkably, only instruments measuring science learning environments with regard to this aspect were found.

[18] ESOL stands for teaching English to Speakers of Other Languages.

Firstly, three instruments point to the integration of reality in science activities in a general way by referring to the use of reality in science as a school subject. One instrument evaluates the connection of science with the outdoor classroom[19] [15] ('How do you use the outdoor classroom across the curriculum?' and 'How does the environment integrate into your teaching of science?'). Two instruments connect science to real-world contexts [31 and 37] and other disciplines [31 and 37]. While referring to 'other disciplines' is not directly linked to reality, one of these instruments [31] includes 'real-world contexts' and 'other disciplines' together in one and the same item ('appropriate connections were made to other areas of mathematics/science, to other disciplines, and/or to real-world contexts'). The same instrument – with the *mathematics/science content* dimension [31] – points to the significance and relevance of mathematics/science (e.g. 'the mathematics/science content was significant and worthwhile'), but it is not clear whether this points to a connection with reality.

While using terms like 'science content', 'science curriculum' and 'teaching of science', the instruments of this first group remain rather vague in that it is not clear in which way reality is incorporated into school science activities.

Nevertheless, six instruments are more specific about this object of study, in two different ways. In a first group, two of the previously discussed instruments can be situated as they determine how often students 'work on solving a real-world problem' [37] and 'In this class, I think about interesting real-life problems' [3]. In a second group, five instruments [1, 6, 9, 14 and 16] assess whether real-life situations or phenomena are directly linked to scientific concepts or principles. The *engaging students with phenomena* cluster [6] stipulates as a condition that the real phenomena are capable of explaining science concepts and the *teacher connected real-life events to teaching math and science principles* scale [14] evaluates the explicit link between the reality brought into the classroom with the scientific principle(s) one wants to teach (e.g. 'teacher drew links between theory and real-life application of math and science principles'). In the latter, the connection with reality is concretised to the application of concepts or skills learned in real life. This also applies to the *elaborate* scale [1], the *student role* scale [16] and a *scientific understanding* scale [9]. These three scales explicitly deal with the application in 'everyday' situations.

Similarly, the *personal relevance* scale [4], which we already discussed before, has, in addition to a general item (e.g. 'students learn about the world inside and outside of school'), a specific item ('new learning relates to experiences or questions about the world inside and outside of school'). Referring to experiences or questions can be seen as another way of being specific about the object of study.

[19] According to Eick (2011), the outdoor classroom can be conceptualised as outdoor experiences in nature, the wider environment.

Science as inquiry

The component *science as inquiry* links strongly with *connection with reality*, but deserves special attention, as inquiry-based learning environments are studied a great deal using scales, items and questions specifically created for that purpose.

In 11 instruments, the integration of reality is perceived as the incorporation of science activities as conducted by 'real' scientists. This can be explained since scientific knowledge is often seen as the result of inquiry; inquiry as such is called 'real' science (Baxter, Jenkins, Southerland, & Wilson, 2004; Taylor, Fraser, & Fisher, 1997). Nine instruments have a scientific inquiry scale or dimension [9, 10, 26, 27, 29, 30, 33, 31 and 35]. The *scientific inquiry* scale of two classroom observation guidelines [9 and 10] is the most explicit in its conceptualisation of inquiry as 'the extent to which students engage in investigation/experimentation [9] or scientific inquiry [10]'. Two other instruments have no specific 'inquiry' scale, but pose several questions with regard to the amount of time the teacher spends on inquiry, NOS, experimental design, processes and skills [8] and with regard to the strategies the teacher uses to teach science (with a focus on scientific inquiry and understanding) [34].

Although different instruments [5, 6, 7, 14, 31 and 37] do not mention the concept of 'inquiry', they do register how often students participate in designing and conducting science experiments or investigations or watch their teacher do a science experiment. Several items (e.g. assessing how often students participate in 'looking at something like the weather or a plant growing and writing down what I see' [5]) and five scales/dimensions (the *developing and using scientific ideas and skills during instruction* scale [6], the *design* dimension [31], the *implementation* dimension [31], the *teacher highlighting the process of science and mathematics* scale [14] and the *critical thinking* scale [14]) exist to measure the authenticity of the process of science conducted in the classroom. Remarkably, while the concept of 'hands-on' activities is often used in the literature on science education, only two instruments [7 and 37] have an item measuring how often students participate in them (e.g. asking teachers how often students 'engage in hands-on activities' [37]). One of these [37] and three other instruments [10, 12 and 31] explicitly recognise that the process of science is not only associated with doing experiments or investigations, but also with higher order thinking. Such association should not cause any surprise as combining investigations and experiments with reflections about what one is doing is strongly emphasised in the literature (not only 'hands-on' but also 'minds-on' activities) (e.g. Hodson, 2008). Characteristic for 'real scientists' is their reflection on theory, hypotheses and findings stemming from their research. Different items (e.g. how often students 'write reflections in a notebook or journal') and three scales (a scientific inquiry scale [10], the *teacher highlighting the process of science and mathematics* scale [14] and the *critical thinking* scale [14]) elaborate on this. One instrument [3] goes even further by focusing on the broader framework in which scientific knowledge develops, especially with the *uncertainty* scale that measures the extent to which students get opportunities to experience that scientific knowledge is evolving and culturally and socially determined.

Interesting activities

While it is assumed that the connection with reality makes S&T activities interesting for students, there are other ways to make activities interesting. The concept of 'student engagement' is used in this context, which is also reflected in the instruments that evaluate the learning environment.

Five instruments [1, 6, 31, 35 and 36] address the engagement or active participation of students by encouraging their questions, generating their interest in the activities to be taught and in what follows (e.g. 'students were intellectually encouraged with important ideas relevant to the focus of the lesson' [31]). The *engage* scale [1] and the *providing a sense of purpose for students and the teacher* scale [6] are elaborated the most (e.g. 'the engage phase raises student interest/motivation'). Another instrument [21] – with the *satisfaction* scale – focuses on students' enjoyment of class work. One instrument [37] appears to determine the frequency by which challenging activities are provided in the classroom by assessing how often students 'design objects within constraints (e.g. egg drop, toothpick bridge, aluminium boats)'.

.

Group work

When examining how the different instruments evaluate group work in the classroom, we noticed that different terms are used to refer to students working together. 'Cooperative' and 'collaborative' learning are the most commonly used concepts. While these terms are often used interchangeably, some authors like Marx et al. (1997) make a clear distinction between 'cooperative' and 'collaborative' learning. According to them, 'cooperative learning' is often highly structured and students are assigned roles, tasks and procedures. The tasks are generally designed to help them assimilate information already presented or to solve problems provided by the teacher. Collaboration, by contrast, is more loosely structured with roles largely negotiated among participants. Cooperation focuses on small groups within the classroom; collaboration involves communities of learners. Nevertheless, the terms 'cooperation' and 'collaboration' are not always used in the instruments or by the developers and researchers discussing those instruments.

The assessment of group work can be categorised in (a) providing opportunities for group work, discourse and negotiation; (b) the suitability of materials and activities; (c) connecting group work to other aspects of the learning environment, such as classroom climate and class management, which promote or inhibit effective group work in class; and (d) digging deeper into effective group work in the classroom. The first category deals with instruments that evaluate whether and how frequently group work in the classroom occurs; instruments that provide measures concerning a further characterisation of this group work belong to the subsequent categories.

Firstly, nine instruments [3, 4, 5, 9, 10, 14, 16, 18 and 37] determine whether and how often group work, discourse and negotiation occur in the classroom without making an explicit connection with the design of the lesson (a). Five scales – the *negotiation* scales [3 and 4], the *scientific discourse* scale [9 and 10], the *an attempt was made to promote discourse and communication of ideas* scale [14], the *classroom culture* scale (subscale: *communicative interactions*) and the *student role* scale [16] – and several items are used to measure this aspect. The *scientific discourse* scale [9 and 10] focuses the most on the quality of group work as this scale evaluates to what extent classroom discourse is developed to create or negotiate shared understandings of science. Only four instruments [5, 16, 22 and 37] have an item that explicitly measures whether students have the opportunity to work in small groups. In the same item ('students work and discuss ideas in small cooperative groups or pairs as well as in whole-group meetings'), one of them [16] points to working in whole groups (with the whole class). It can be assumed this item refers to both cooperative as well as collaborative group work, although these concepts are not used. Only one instrument [37] explicitly assesses how often students 'work in cooperative learning groups'.

Secondly, six instruments [1, 14, 20, 22, 31 and 37] measure whether the materials and activities provided by the teacher are suitable for group work (b). Three scales and one dimension – the *learning context* scales [20], the *activities and tasks* scales [20], the *engage* scale [1] and the *design* dimension [31] – address whether the design of the lesson was appropriate for group work. Student discussion and discourse appear to be enhanced by the 'right' organisational forms (project work rather than lectures; arrangement of seating), provided topics and materials. One instrument refers to project work [14] – with two items ('teacher emphasised learning through classroom projects rather than through lecture' and 'students are encouraged to research, present, and assess project work as a group') – and explicitly uses the term of cooperative learning in another item. Two other instruments [22 and 31] refer to collaborative group work in their operationalisation.

Thirdly, a few instruments evaluate other aspects of the learning environment connected to group work in the classroom (c). One instrument [16] – with the *Management Features* scale – connects class management with effective group work. It should be mentioned that one of the two items in this scale that are relevant for cooperative group work ('students take responsibility for maintaining cooperative work environment') is also classified in the *level of initiative* component. Another instrument [31] – via the *classroom culture* dimension – appears to attach value to a positive atmosphere to result in effective group work (e.g. 'there was a climate of respect for students' ideas, questions and contributions'). One of the items of this dimension reflects the importance of collaborative working relationships ('interactions reflected collaborative working relationships between teacher and students'). One other instrument [21] – with the scales *cohesiveness*, *friction* and *competitiveness* – also measures the quality of relationships in the classroom, but does not explicitly connect classroom climate with group work. Eick (2011) finally asks the question, 'how does your use of the outdoor classroom and related activities build classroom community among your students?' [15], to dig deeper into classroom climate, but this question makes no clear connection with group work either.

Finally, some data collection instruments [20, 22 and 25] go a step further by determining the catalysts that make group work in the classroom effective (d). Two of them [20 and 25] dig deeper into the way students interact. One instrument [20] consists of 11 *group interaction* scales to capture the quality of the interactions. It also elaborates on the promotion of competences relevant to working effectively in groups using some of the *role of adults* scales. Since the teacher has an important role in stimulating these competences, these scales could also be classified in the component *teachers' interventions during classroom activities*. Another instrument [22] does not explicitly focus on the interactions or promotion of competences, but more broadly on the use of group work, how groups are organised in science lessons and what teachers see as the most important outcomes of group work in science. Teachers were also asked to describe an example of their most successful experiences of group work in science lessons and to record what might help or hinder students' progress in engaging with group work. None of the three instruments belonging to this category use the terms cooperative or collaborative learning in their items.

An instrument referring to *collaboration and support* as an active learning strategy [35] could not be placed in one of the above categories because more details or information on its content were not provided.

Level of initiative

Eleven instruments measure the amount of freedom the teacher gives students during S&T activities. While some scales, items and questions evaluate whether students' interests and ideas are an important starting point during the lesson, others refer to students' autonomy in the execution of the task. Others go further by evaluating the participation of students in determining and organising the S&T activities in the classroom. Two instruments evaluate different aspects of the learning environment starting from an autonomy perspective.

A first group comprises those instruments [1 and 6] that refer to the teacher who takes students' questions and ideas into consideration. Because of the connection between lesson activities and students' lives, one could initially argue that the scales evaluating this aspect should have been classified under the component *anticipating students' prior knowledge and backgrounds.* Still, they fit better in the *level of initiative* component as students are actively asked for their interests and/or their interests and questions are taken into consideration during classroom activities, too – and so after the teacher has – eventually – prepared the activities. More specifically, the *explore* scale [1] (with the item 'learning activities in the exploration phase are student-centred', meaning that 'when appropriate, teacher questions evoke the learners' ideas and/or generate new questions from students. Student inquiry may involve student questioning, manipulating objects, developing inquiry skills (as appropriate) and developing abstract ideas') and the *taking account of student ideas to inform instruction* cluster [6] are relevant scales in these context.

We can place the instruments [3, 5 and 37] that measure the freedom students receive in executing a task in a second group. Instruments belonging to this group determine how often students participate in conducting their own investigations [5 and 37] and how much time they spend on an activity [3] (the *autonomy* scale).

A third group of instruments [4, 16, 18 and 31] assesses students' freedom in even more depth because their contribution to the determination and management of classroom activities is measured. Four scales with some of their items – the *critical voice* scale [4], the *shared control* scale [4], the *lesson design and implementation* scale [18] (item: 'the focus and direction of the lesson was often determined by ideas originating with students') and the *management features* scale [16] – and one separate item (determines how often students 'participate in student-led discussions') [37] evaluate whether students have a say in the sort of classroom activities that are conducted and how these are organised. Of these, the *critical voice* scale [4] and the *shared control* scale [4] are theoretically well founded as the instrument to which these scales belong starts from a critical theory perspective (Taylor et al., 1997). In this perspective, scientific knowledge is not fixed, but results from continuous human inquiry and must be validated against community norms. While the *management features* scale [16] integrates items referring to classroom management as co-determined by both students (e.g. 'students take responsibility for maintaining cooperative work environment') and the teacher (e.g. 'teacher consistently monitors students' behaviour, acknowledges appropriate behaviour, and applies agreed-on consequences'), another instrument [31] only focuses on the teacher when dealing with class management (with the items 'the teacher's classroom management style/strategies enhance the quality of the lesson' and 'the design of the lesson reflected careful planning and organisation'). Even though one could situate the item of [18] in the second group of level of initiative, it was categorised in this group because of its strong terms, such as 'focus' and 'direction', which may indicate students' contribution to the determination and management of classroom activities.

While the above-mentioned categories contain instruments that measure students' freedom as only one aspect of the learning environment, two instruments more holistically evaluate different aspects of the learning environment starting from an autonomy perspective. In the DASTT-C [17], teachers have to make a picture of themselves and their students while teaching science. Teachers are rated on different dimensions of giving autonomy to students on the basis of their pictures: students' management of their own learning, flexibility of the curriculum to students' interests, the role of the teacher as a coach rather than a provider of information, the teacher's focus on students' questions rather than on being the initiator of activities and the acknowledgement of students' input. The STIR [19] consists of scales that measure the level of autonomy students get during each stage of the inquiry process. This instrument identifies and classifies inquiry activities for each of the five discerned features of classroom inquiry[20] and their variations based on the amount of learner self-direction and direction from materials (NRC, 2000, as cited in Bodzin & Beerer, 2003). The underlying idea is that the five features of classroom inquiry can be incorporated into the science classroom in a highly structured format, with teachers and/ or materials directing students towards known outcomes or they may take the form of open-ended investigations that are learner centred.

Finally, it is rather unclear how some instruments assess students' autonomy. While we assume that the *student-centeredness* scales of two instruments [3 and 7] can be related to students' freedom in the classroom, we do not know the exact meaning of 'student-centeredness' in these scales. We know that one *student-centeredness* scale [3] – or in reverse scoring *teacher expectations* – aims to measure teacher expectations for different aspects of learning ('In this class, the teacher expects me to remember things I learned in past

[20] The five essential features of inquiry-based teaching are: learners are engaged by scientifically oriented questions; learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions; learners formulate explanations from evidence to address scientifically oriented questions; learners evaluate their explanations in light of alternate explanations, particularly those reflecting scientific understanding; learners communicate and justify their proposed explanations (NRC, 2000, p. 4, as cited in Bodzin & Beerer, 2003).

lessons'), but it is not clear whether and how this addresses students' initiative. Similarly, the item 'the lesson was modified as needed based on teacher questioning or other student assessments' in the *implementation* dimension [31] is rather general in nature.

Fostering understanding

Different instruments measure whether the teacher fosters students' understanding in science activities. In addition, a few instruments also focus on some other conditions, apart from the provided science activities, which can contribute to students' understanding.

Firstly, instruments assess whether science activities in the classroom contribute to students' understanding. While some scales, items and questions [10, 12, 16, 18, 30, 31 and 34] directly and generally address activities that promote conceptual understanding (e.g. the *scientific understanding* scale [10] reflects on 'the extent to which knowledge is treated in a shallow and superficial manner'), others are more specific by pointing to students actively working using a variety of means to represent phenomena [18] or working on models or simulations [37], rather than doing more superficial activities like reading a science (text)book in class [5 and 37], answering textbook/worksheet questions/assignments [37], giving or listening to formal presentations [31 and 37] or following specific instructions in an activity or investigation [37] that may result in rote learning. Five scales – the *scientific understanding* scale [10], the *scientific inquiry* scale [12], the *content* scale [18], the *content features* scale [16] and the *Practice in Scientific Understanding* scale [30], several items [5, 31 and 37] (e.g. 'adequate time and structure were provided for wrap-up' [31]) and questions [34] (e.g. 'I'd like to know about the strategies that you use to teach science – with a focus on scientific understanding and inquiry') exist to measure aspects related to fostering students' understanding.

Secondly, some instruments refer to another aspect, separate from the provided science activities, which is important to take into account when fostering students' understanding. Four instruments [9, 11, 12 and 13] measure the extent to which students' home language is used to enhance understanding in regular (non-bilingual) classrooms - two of them use the students' *home language* scale [9 and 11]. Two instruments [12 and 13] also evaluate the students' home culture.

Even though one instrument [26] appears to assess the stimulation of students' understanding, it is not clear on the basis of the article found how this is operationalised.

Use of new technology

The use of new technology, meaning information technology (digital media) and not design, is rarely assessed. Four instruments [1, 14, 35 and 37] evaluate the availability, use or satisfaction of technology in the classroom (e.g. 'teacher required students to use computer or calculations' [14]). The use of technology is more indirectly evaluated in one of them [1] – via the *explain* scale – since it is viewed as one possible approach to explain and illustrate concepts or skills.

Teachers' interventions during classroom activities

While the previous components mainly dealt with system characteristics of the learning environment, this component deals with the teacher's actions during class activities themselves and with the way he or she intervenes and interacts with students during those activities. Eleven instruments explore the role of the teacher during S&T activities. Three different ways of assessment can be identified from among these.

Firstly, four instruments [1, 2, 18 and 31] assess whether and/or how teachers encourage students to approach scientific activities, by giving directions, responding to students and encouraging them to find answers on their own (the *explore* scale [1]) (in this scale the 'teacher gives directions, responds to students, and encourages students to find answers on their own'), by encouraging them to collect and analyse data [2], by acting as a resource person, by supporting and enhancing investigations (*classroom culture* scale (student-teacher relationships) [18] (e.g. 'the teacher acted as a resource person, working to support and enhance student investigations') and by encouraging and valuing the active participation of all students [31].

Secondly, six instruments [1, 2, 5, 16, 31 and 37] evaluate the opportunities teachers provide for students to reflect, to express their ideas, activities and findings. Five of them [1, 2, 5, 16 and 37] particularly assess whether teachers ask students to explain their topics under investigation and their findings (e.g. asking for students' perceptions of the frequency of participation in 'writing or giving an explanation of something I'm studying in science' [5]). The *explain* scale [1] and the *teacher role* scale [16] – together with some items from another instrument [37] (e.g. teachers are asked how often they 'require students to supply evidence to support their claims') – are elaborated on the most concerning this aspect of teacher-student interactions. Unsurprisingly, some instruments [1, 16 and 37] explicitly connect these questions for explanation with the development of concepts in students (e.g. 'teacher facilitates students' construction of science conceptions by contrasting ideas, encouraging discussion, asking for applications, and/or modelling cognitive processes' [16]). One other instrument [31] assesses whether teachers pose questions that encourage students' conceptual understanding/problem-solving (item: 'the teacher's questioning strategies were likely to enhance the development of student conceptual understanding/problem solving').

Thirdly, three instruments [7, 8 and 17] use the concept of 'teaching style', which in the literature often refers to the way in which the teacher intervenes during classroom activities (see, for example, Laevers & Heylen, 2013), although the concept is rather vaguely used (e.g. 'I made some general changes to the instructional strategies and my teaching style' [7]) in the found instruments. The DASTT-C [17] measures – with regard to several aspects of the learning environment – the extent to which lessons are directed by the teacher.

One instrument [26] has a scale or dimension referring to interaction (attitude concerning interaction), but no details could be retrieved.

Teachers' content knowledge

In seven instruments [9, 10, 11, 30, 31, 33 and 35], teachers' scientific content knowledge – no instrument measures teachers' technological content knowledge – is evaluated. Five instruments [9, 10, 11, 30 and 33] assess the accuracy and comprehension of teachers' mastery of the science content provided in the lesson via the same scale, the *teachers' knowledge of science content* scale. One instrument [31] not only evaluates teachers' scientific content knowledge, it also evaluates whether 'the teacher appeared confident in his/her ability to teach mathematics/science'.

Evaluating students' understanding

When determining the quality of the learning environment, six instruments assess whether the teacher evaluates students' understanding and the types of evaluation that are used. Two ways to classify assessment can be determined.

Firstly, some instruments clearly refer to the concepts of 'summative' and 'formative' evaluation. While summative evaluation assesses students' knowledge and skills after instruction on (a) certain topic(s) is provided, formative types of evaluation are embedded in the learning material(s), meaning that evaluation takes place during instruction. The first form of assessment is evaluated via the *evaluation* scale [1], the second via the *examining assessment of progress* cluster [6], the *explore* scale [1] and via some questions from another instrument [37] (e.g. asking how often teachers 'use assessment to find out what students know before or during a unit').

Secondly, different instruments assess whether the type of evaluation chosen by the teacher is in line with the learning goals. Four instruments [6, 17, 31 and 37] use alternative forms of assessment to evaluate whether students show a profound understanding instead of a superficial one. These alternative forms are operationalised as being consistent with investigative mathematics/science (the *design* dimension [31]), requiring open-ended responses [37], engaging students in performance tasks [37], requiring application of the science ideas and not allowing students a trivial way of responding (the *examining assessment of progress* cluster [6]). One instrument [17] is less clear about these alternative ways of assessment, apart from stating that they measure students' learning and knowledge in a holistic way, in contrast to tests that check for understanding of important concepts or focus on scientific content knowledge.

One instrument [35] contains questions about student assessment, but these are not given in the article in which the instrument was used.

Assessing project-based learning environments with the retrieved instruments

In the previous section, we discussed 11 components that serve as categories in which each of the retrieved instruments with one or more operationalisations could be placed. As may become clear, some of these components fit in more than others with the discerned aspects of project-based learning environments as discussed in the introduction. In order to answer our second research question ('Which scales, items and questions fit in with the aspects of project-based learning environments?'), we are particularly interested in those scales, items and questions belonging to the components coming close to these features. In the following section, we take the aspects of project-based S&T learning environments, which we discussed in the introduction, as our starting point.

Nevertheless, some other components retrieved in the bottom-up analysis – and their operationalisation – cannot be linked directly to the aspects of project-based learning environments. Still, we should be aware of the possible importance of scales, items and questions categorised under these components for the evaluation of project-based learning environments since they can provide new insights into their effective and less effective aspects. In the end, this can contribute to the further building of theories with regard to these learning settings.

Challenging problems

Providing a challenging problem in project-based learning environments is a recurring theme in the literature (Jones et al., 1997; Liljeström et al., 2013; Luera & Otto, 2005; Marx et al., 1994; Pucel, 1992). One question was found that assesses this feature by determining how often students 'design objects within constraints (e.g. egg drop, toothpick bridge, aluminium boats)' [37] (see *Interesting Activities*).

Authenticity

The connection of the given problem with students' everyday lives is typical of project-based learning environments (Harel & Papert, 1991; Kafai & Resnick, 1996; Rogers et al., 2011; Thomas, 2000). While the *personal relevance* scale [4] can be relevant, as it measures the relevance of school science to students' everyday lives, the items of this scale do not point to the given problem, but to experiences, questions or the world in general (see 'connection with reality'). Two items, 'In this class, I think about interesting real-life problems'[3] and one that determines how often students 'work on solving a real-world problem' [37], overcome this difficulty, although it is not clear what exactly is meant with 'real-life' and 'real-world'. A problem taken out of reality can refer to a problem known by the students, connected to their everyday life, but it can also refer to a relatively new real problem, with which they are unfamiliar.

The authenticity of the given problem in project-based learning environments also refers to the fact that problems are characterised by different elements that should be disentangled and for which knowledge of different subjects (S&T, but others as well) is required (Barak & Doppelt, 2000; Barak & Raz, 2000). This differs from the perception of the use of reality in the *elaborate* scale [1], a *scientific understanding* scale [9], the *student role* scale [16], the *engaging students with phenomena* cluster [6] and the *teacher connected real-life events to teaching math and science principles* scale [14] since the explicit link between reality and science concepts or scientific principles forms an indicator to evaluate the connection with reality in these scales. General learning goals in project-based learning environments can be considered in advance in the light of the competencies a given problem should foster (Thomas, 2000), but this specific link between reality and S&T concepts and principles is usually not explicit.

Rather than the relevance of the given problem, the process of scientific inquiry is the main link with reality in most of the instruments. As outlined in our theoretical framework, 'real' inquiry activities – and so their corresponding items, scales and questions – can also be important in project-based learning environments.

Even though project-based activities can be carried out without the *use of new technology* (Marx et al., 1997), giving the opportunity to use new technologies while approaching the given problem can make the environment even more authentic to students. Instruments can enable them to carry out genuine inquiry. Four instruments [1, 14, 35 and 37] can be used to evaluate the availability, use or satisfaction of technology in the classroom.

Students working autonomously

Students get a considerable degree of initiative in project-based learning environments (Holubova, 2008; Thomas, 2000). This initiative is not limited to taking students' ideas and interests into consideration and allowing them freedom in their approach to the given problem; students are also involved in the determination and management of classroom activities (Rogers et al., 2011). Therefore, instruments, scales, items and questions that are more demanding than others with regard to the provision of initiative to students (see '*the level of initiative*') are useful. The *critical voice* scale [4], the *student negotiation* scale [4], the *lesson design and implementation* scale [18] and the *management features* scale [16] are particularly relevant. Questioning how often students 'participate in student-led discussions' [17] is also interesting in this context. The DASTT-C [17] and the STIR [19] are valuable instruments to evaluate different aspects of the learning environment starting from this autonomy perspective.

The active role of the teacher

Along with the considerable degree of autonomy goes the role of the teacher as a guide, a partner in the learning process (Barth, 1972) and not as a mere provider of information. To this day, researchers are still exploring the teacher's role in S&T activities to get insight into the effective ways of intervening in classroom interactions. To gain more clarity with regard to this question, it is therefore important to measure different aspects of the teacher's role. Because of their perception of the teacher as a coach in the learning process, the retrieved scales and items evaluating teachers' interventions during classroom activities are all interesting to use in research on project-based learning environments. The *explore* scale [1], the *classroom culture* scale (*teacher-student relationships*) [18] and some separate items [2 and 31] can be used to measure the support provided by the teacher while students go through the process of design and inquiry. The *explain* scale [1], the *teacher role* scale [16] and some items [31 and 37] provide good measures to assess whether the teacher encourages students to reflect by explaining their ideas and concepts.

Working collaboratively

Students have a lot of autonomy and the amount of guidance from the teacher is rather limited in project-based learning environments. This is covered by other ways to organise class work, e.g. by letting students work together while approaching the given problem. Characteristic for project-based learning environments is that students work collaboratively in a larger community (Liljeström et al., 2013), rather than in small, predetermined groups (cooperative group work). However, only a few of the retrieved instruments explicitly use the term 'collaborative' group work/learning in their operationalisation. Assuming that their conceptualisation of this term corresponds with that of Marx et al. (1997) (see 'group work'), these instruments are initially interesting to consider in order to determine the quality of project-based learning environments. Two instruments [22 and 31] – the latter via the *design* dimension – provide ways to evaluate whether the design of the lesson, the materials and the activities allow a collaborative approach to learning. Via the *classroom culture* dimension of the last instrument [31], it is also possible to determine whether the atmosphere in the classroom is favourable for collaborative group work. While no instrument explicitly provides a way to determine whether and how often collaborative group work occurs in the classroom, one instrument [16] determines whether students, in addition to working in small cooperative groups or pairs, work in whole group meetings, which is typical of collaborative group work (Marx et al., 1997). We do not know whether those instruments that dig deeper into the effectiveness of group work evaluate collaborative or cooperative group work. Although project-based learning environments contain features of collaborative learning, scales, items and questions referring to cooperative group work can also be useful in the determination of their quality.

Deep understanding

In the end, project-based learning environments aim to foster understanding in students instead of rote learning (Rivet & Krajcik, 2008). While a lot of scales, items and questions [10, 12, 16, 18, 30, 31 and 34] directly and generally address activities that promote conceptual understanding or higher order thinking, these are rather unspecific in their description of what those activities should encompass. Others [5, 18, 31 and 37] precisely describe those activities and evaluate whether or how frequently they take place in the classroom; however, these lists of items are not exhaustive and cannot capture all activities contributing to a deep level of understanding in students in project-based learning environments.

With regard to the assessment forms used by the teacher to evaluate whether students show a deep level of understanding, the bottom-up analysis clearly showed that four instruments [6, 17, 31 and 37] use alternative forms of assessment – consistent with investigative mathematics/science (the *design* dimension [31]), requiring open-ended responses [37], engaging students in performance tasks [37] and requiring application of the science ideas and not allowing students to respond in a trivial way (the *examining assessment of progress* cluster [6]). Since we assume project-based learning environments function according to the open framework model (Laevers, 2011; Schweinhart & Weikart, 1997), it is not completely clear which learning content is used in these settings to bring the project to a favourable conclusion. A consequence of the 'emergent' curriculum is that not all students will acquire the same competencies and certainly not at the same pace. We therefore have to question whether it is (1) possible to create a test and (2) expedient for the teacher to use one and the same test to assess students' understanding. Maybe project-based learning environments require even more 'alternative' ways of assessment. Nevertheless, the teacher can observe whether and what students are learning during the learning process. The presence of these formative types of evaluation can be assessed using the *examining assessment of progress* cluster [6], the *explore* scale [1] and some questions [37].

Conclusions and discussion

This review reports empirical research on instruments that measure the quality of project-based learning environments with a focus on S&T in elementary schools. The research questions were twofold: (1) Which aspects are measured by the existing instruments to gain insight into the quality of S&T elementary learning environments and how are these operationalised? and (2) Which scales, items and questions fit in with the aspects of project-based learning environments that were outlined above?

To answer the first research question, we conducted a horizontal analysis which clearly showed that different aspects are evaluated in the reviewed instruments. We found a total of 11 components: anticipating students' prior knowledge and backgrounds, connection with reality, interesting activities, group work, the level of initiative, teachers' interventions, teachers' content knowledge, use of new technology, fostering understanding and evaluating students' understanding. Of these components, *anticipating students' prior knowledge and backgrounds*, the *connection* of the provided materials and activities *with reality* (together with *science as inquiry*), *group work* and the *level of initiative* students receive during classroom activities are evaluated by the largest number of instruments. Even more importantly, these components differ greatly in their operationalisation. These differences are related to both the object of measurement and the extent to which the used concepts are clarified in the particular instruments, particularly for the components *connection with reality*, *group work* and *level of initiative.*

With regard to the object of measurement, we first found a varying level of specificity in scales, items and questions. While some scales, items and questions more generally connect the 'science curriculum' or 'science content' to real-world contexts, other disciplines or the outdoor classroom, others are more specific and relate the problem provided by the teacher to reality or even provide a direct link between the scientific concept or principle and reality. The *personal relevance* scale [4] that we categorised in *anticipating students' prior knowledge and backgrounds* has both a general and a specific item, the latter relating to experiences and questions. Furthermore, a large group of instruments evaluate the incorporation of science activities as conducted by real scientists or science as inquiry. Dependent on their goals, it is therefore important that researchers are aware of these existing levels of specificity when choosing a particular scale, item or question. A second conclusion with regard to the object of measurement relates to differences in requirements reflected in scales, items and questions. This is the case for *level of initiative*, in which some scales, items and questions take students' interests and questions into account when making an evaluation, while others focus on their autonomy in doing a task, and still others explore whether students can participate in the determination and management of classroom activities. When investigating the students' level of initiative in the classroom, educators should therefore bear these requirements for students' level of initiative reflected in scales, items and questions in mind. Thirdly, it was also found that a variety of aspects related to one and the same concept were evaluated, in particular for the component *group work*. This component is assessed by evaluating the opportunities for group work, discourse and negotiation; the suitability of materials and activities for group work; other aspects of the learning environment (such as classroom climate and class management) related to group work; and by digging deeper into effective group work. With regard to this component, researchers appear to have a spectrum of possibilities when selecting those scales, items and/or questions that fit in with their research questions. Nevertheless, another problem that hinders researchers arises here.

That brings us to the clarification of the used concepts, which is – in addition to the object of measurement – responsible for the differences in operationalisation of the most frequently assessed components. While a few instruments use the terms cooperative and collaborative group work – two ways in which group work can be seen – it is not always clear how these are conceptualised in the different instruments. A similar problem exists for the measurement of the connection between school science, a given problem or scientific principle or concept, and the real world or real life. The link with reality can refer to a reality known by the students, connected to their everyday life, but it can also refer to relatively new real-world or real-life contexts, with which they are unfamiliar. However, not knowing the conceptualisation of different terms makes it difficult for researchers to select previously used instruments, scales, items and questions for their own research. Therefore, with this review, we want to call on researchers to be more clear about the specific instruments, scales, items and questions they use, for example, by providing a short guidebook or explanation in the appendix section of their articles. This may result in colleagues in the field being more willing to adopt particular instruments, scales, items and questions for their own research. Using instruments to evaluate the quality of learning environments in S&T education repeatedly can result in the optimisation of their reliability and validity in the long run.

As a consequence of the differences in the object of measurement and the clarification of the used concepts, some scales, items and questions, depending on their operationalisation, were found to fit in better than others with the aspects of project-based learning environments (providing a challenging problem, authenticity, students working autonomously, working collaboratively, the active role of the teacher and deep understanding) (second research question). The differences in requirements in the *level of initiative* component allowed us to select appropriate instruments and scales for the similar aspect in project-based learning environments. Other components, like *teachers' interventions*, *fostering understanding* and *group work*, provided us with the evaluation of a variety of aspects. *Teachers' interventions* gave us the necessary scales and items to broadly explore what constitutes the active role of the teacher in project-based learning environments. That was less the case for deep level of understanding, as not many of the retrieved scales, items and questions point to specific activities that *foster students' understanding*. While a variety of aspects are evaluated in group work, more scales, items and questions could have been selected to measure collaborative group work, had the concepts of 'cooperative' and 'collaborative' group work been used and clarified. Similarly, more clarity about the term 'reality' would have resulted in more useful scales, items and questions for the evaluation of the authenticity aspect. More specificity about the aspects of the curriculum in which that reality and students' everyday lives are integrated in the science curriculum also appeared to be important. Still, the result of only one survey question to measure whether the provided problem is challenging is not related to differences in the object of measurement and the clarification of the used concepts.

Although it was possible to select some scales, items and questions in order to assess each of the aspects of project-based learning environments, most of the retrieved instruments are related to science learning environments, and not to S&T or to project-based learning environments. This became especially apparent for those scales, items and questions that measure *science as inquiry*. While these are also interesting for use in project-based learning environments, one should be aware of the importance of measuring the quality of the design process as well, given that design education is a distinct discipline with its own characteristics (Jones, Buntting, & de Vries, 2013) and is often integrated into project-based learning environments. However, as mainly instruments in the field of science education were found and only one [1] appears to be appropriate for research on design education, too, the horizontal analysis provided no evidence for the existence of instruments evaluating the quality of the design process. This outcome raises the expectation that research in design education and project-based learning environments is not often conducted with instruments specifically developed for that purpose. A lot of research is qualitative and descriptive in nature. While we do not deny the value of this type of research, and we do recognise that research in design, and particularly in project-based education, is relatively young (Thomas, 2000), we think it is important to create a balance. Taking a much closer look should make it possible to provide more insight into what works and what doesn't in project-based learning environments. This need is high as Thomas (2000) already pointed out in his review of research on project-based learning 15 years ago, specifying the importance of more research documenting the effective aspects of project-based learning environments. This is particularly important if we want more schools and teachers – who are increasingly pressured to attain the standards with their students and who have to account for their practices (Thomas, 2000) – to use project-based learning environments. A good starting point is trying out the existing scales, items and questions as they are used for the closely related field of science education. In a next step, these scales, items and questions can be adapted to the specific context of design and project-based S&T education.

Although we could draw some interesting conclusions with regard to instruments in project-based learning environments, there are some limitations of the way this review is conducted. A first limitation might be the exclusion of book chapters and conference papers in our data-set. Secondly, we mainly analysed the instruments as used by the authors of the retrieved articles. We only consulted the original source if the article provided insufficient information on an instrument for which the original source was given. A third limitation is related to the inevitable fact that we had to make an interpretation when situating scales, items or questions in a certain category (characterised by features recognised in the instruments). In doing so, we tried to be as objective as possible by only relying on the information given by the authors in the retrieved articles (or in some cases, in the articles to which the authors refer to). In some instances, only a sample item of a scale was available (e.g. 'In this class I think about interesting life problems' of the Prior Knowledge scale of the CLES [3]); then, we could only rely on this sample item for situating the scale with its item in a specific component category. Even if these limitations are taken into account, this systemic review enabled us to get a concise overview of instruments used in studies over the last 15 years that research the effectiveness of S&T learning environments. Future studies in these domains can benefit from this overview when searching for an instrument, scale, item or question to measure one or more particular aspects of the learning environment. Even though the particular goal in this article was to select particular scales, items and questions for mapping the quality of project-based learning environments, all necessary information is provided to select specific scales, items and questions for any learning environment in which S&T activities are conducted. Finally, this review encourages researchers in the field to compose new instruments with scales of which the operationalisation fits best with their goals. Testing can optimise our insight into the working of effective S&T learning environments.

Study 2: The effectiveness of a project-based S&T learning environment for pupils' growth in engagement: Investigating explanatory class factors

Submitted

Keywords: primary school, science & technology (S&T) education, pupil engagement, teacher attitudes towards S&T (and its teaching), teacher-pupil and pupil-pupil interactions

Abstract

This study investigates the effect of the project-based science and technology (S&T) learning environment Village@School on pupils' growth in engagement and potential class factors explaining this growth (teachers' attitudes towards S&T and its teaching and interactions in the classroom). The investigation took place over a 2-year period in 18 primary schools in Belgium (Flanders) and The Netherlands. Data was primarily collected via classroom observations. The findings reveal that (a) pupils grew in their engagement throughout the implementation of the project Village@School and (b) the growth in engagement as measured before and after the project was positively related to the sensitivity of the teacher throughout the project, but negatively to a positive class climate and an emphasis on a deep level of content understanding when other dimensions of interactions remain constant. The results of this study provide insight into the conditions for success of project-based S&T learning environments for pupils in primary school.

Introduction

Internationally, several authors have reported a decline in enthusiasm for science (Mant et al., 2007). Pell & Jarvis (2001) noted that this decline appears to begin towards the end of primary school. Murphy and Beggs (2003) remarked that in the final years of primary school science is frequently being taught as a 'body of knowledge'. Teachers often teach science in this way because they are convinced that otherwise the standards cannot be reached.

However, open-ended learning environments in which students can experience a lot of autonomy, like self-regulated (Boekaerts, 2002), inquiry, design- and project-based learning environments (Barak & Raz, 2000), are very popular in the literature concerning S&T education. As students can design and implement their own inquiries, they are permitted a lot of freedom, thereby engaging in and achieving ownership of their learning (Newmann, Wehlage, & Lamborn, 1992). From this viewpoint students are actively involved in doing science, which is different from passively learning about science (Meyer & Crawford, 2011). Open-ended learning environments are beneficial for students' outcomes; including their affective outcomes, such as motivation (Barak & Raz, 2000; Barron et al., 1998; Doppelt, 2003; Liu & Hsiao, 2002; Westwood, 2006) and engagement (Kaldi et al., 2011; Mant et al., 2007; Wurdinger, Haar, Hugg, & Bezon, 2007).

Primary school pupils are not always used to such open-ended learning environments. They experience a lot of autonomy compared to the traditional approaches like whole-class teaching in which teachers tend to provide their students with full instructions and minimal opportunities to design and investigate (Gott & Duggan, 1996). As a consequence, students have to learn how to move themselves in such open learning environments (Windschitl, 2003). It is not evident for them to perform independent investigative work which requires the ability to identify a problem, formulate a hypothesis, interpret experimental results, and evaluate conclusions (Champagne, Gunstone, & Klopfer, 1985; Tamir, Stavy, & Ratner, 1998). Therefore, we may expect that some conditions should be fulfilled to help pupils to cope with the challenges of an open framework and as a result grow in their engagement in S&T activities.

These reflections result in the two-fold aim of this study. Firstly, we want to investigate how pupils' engagement evolves in the course of a challenging S&T project known as Village@School. In this project pupils are challenged to build a miniature site on a standard plate, and are given a lot of autonomy to do so. This may give us some insight into the potential of this project to increase pupils' engagement in S&T in upper primary school. Secondly, we want to gain more insight into the active ingredients that explain pupils' increase in engagement in such challenging S&T activities. There is a need to establish a more fine-tuned view of what works and what does not in these environments, partly to respond to those who criticize the insufficient guidance typical of these approaches (e.g. Kirschner, Sweller, & Clark, 2006). In fact, guiding pupils through inquiry- and design-based learning environments is a real challenge for teachers. A constructivist approach to education has long been encouraged, but in daily classroom practice most of the time teachers still act as providers of new information, standing in front of the class (Galton & MacBeath, 2002; Osborn, McNess, & Broadfoot, 2000) and teaching a subject with particular learning goals in mind. In contrast, in the 'new' open-ended S&T learning environments the curriculum 'emerges' (Laevers, 2011) as they guide pupils through the process of doing inquiry and, at the same time, give them autonomy. Primary school teachers experience a lot of insecurity in doing this, often because they have little affinity with the subjects of S&T themselves and the recommended teaching techniques in those fields. In the literature, the negative attitudes of primary school teachers towards S&T are often described (e.g. Cobern & Loving, 2002) and in the past different professional programmes have been established in order to change them (Osborne & Dillon, 2008; Van Driel et al., 2001).

Theoretical framework

In what follows, we will provide our conceptualisation of pupils' engagement in open-ended S&T learning environments, as well as the factors that may affect this (growth in) engagement: their teacher's attitudes towards S&T (teaching), and the quality of teacher-pupil and pupil-pupil interactions. Empirical evidence from studies about these or similar variables is also summarised. Finally, we will elaborate on the intervention during which the connection between the interactions and pupils' growth in engagement will be investigated.

Pupils' engagement

Linnenbrink and Pintrich (2003) point to the importance of students being cognitively engaged in order to effectively learn and improve their self-efficacy. As an important goal of open-ended learning S&T environments is that students combine 'hands-on' activities with 'heads-in' activities (Papert, 1980), students are given a lot of opportunities for engagement. In such learning environments students do not merely learn theories, but – via the activities and materials provided – they can explore the reality around them and reflect upon it. In the next paragraphs, we will investigate under which conditions pupils' growth in engagement would be higher; but first, we will shed light on our conceptualisation of engagement.

In the literature, three different dimensions of engagement are discerned. The emotional dimension deals with showing interest, giving value, and positive affect; the cognitive dimension with the use of cognitive and metacognitive strategies; and the behavioural dimension with showing effort, persistence, and seeking help (Linnenbrink & Pintrich, 2003). Researchers often use a combination of these dimensions in their conceptualisation of engagement (Fredricks et al., 2004). In this study we use engagement as conceptualised by Laevers (2011), under the notion of 'involvement'. Involvement is characterised by concentration and intense, intrinsically-motivated mental activity (Laevers, 2011). It connects with the 'state of flow' as conceived by Csikszentmihalyi (1988): "Flow is a subjective state of complete involvement, whereby individuals are so involved in an activity that they lose awareness of time and space" (Csikszentmihalyi, 1988, in Fredricks et al., 2004, p. 63). The three dimensions of engagement (Fredricks et al., 2004; Lawson & Lawson, 2013) can be recognised in Laevers' concept of involvement (2011). Involvement is about the 'intensity of mental activity' and – like flow – is not linked to levels of competence. Involvement occurs when a person is operating at the very limits of his capabilities; at whatever level that may be. In what follows, we will not talk about 'involvement' but about 'engagement', as this term is more commonly used to refer to the concept as described above.

Quality of interactions

In open-ended S&T learning environments, interactions are intended to occur in three different ways.

Firstly, interactions with material(s) take place. Students get feedback on their success or failure in building and improving the system they are working on (Barak & Shachar, 2008). According to Wurdinger et al. (2007), a valuable moment in the learning process is when students make mistakes and struggle during the process. They learn from their mistakes and realise that they must re-evaluate their plans and implement them in different ways until they find a solution.

Secondly, in these student-directed learning environments, pupils have a lot of interactions with each other as they work collaboratively to achieve a common goal (Kaldi et al., 2011). They have shared problems, have to come to a consensus on possible solutions, and find a mutually acceptable way of solving the problem (Liljeström et al., 2013). According to Engeström (1992), this process will enhance reflective communication and is the highest form of collaboration. This can foster the clarification of ideas (Kaldi et al., 2011), the internalisation of content knowledge (Cross, 1998), deep learning (Kolodner, 2006b) and positive interdependence (Kaldi et al., 2011). However, reaching the highest form of collaboration does not happen automatically. Wurdinger et al. (2007) showed that some students contributed more than others in their different groups in this type of learning environment. Chanlin (2008) found that although disagreement and conflict among group members sometimes occurred, this reflected a healthy internal growth among group peers. Even when students do not listen to each other when there is disagreement, they reciprocally influence each other's arguments, and exchange views and ideas (Sawyer, 2004).

Thirdly, interactions between students and teachers continue to be important in these open-ended S&T learning environments. Liljeström et al. (2013) found that teachers cannot stay in the background and allow purely student-centered work as described by Kirschner et al. (2006). Teachers are no longer – as in the traditional approach – supposed to give pre-determined aims and guidelines; nor do they need to engage students in experiments of which they already know the correct answers (Mueller, 2011). On the contrary, their role is often described as that of a tutor, a guide and a partner in the learning process (Barth, 1972; Wurdinger et al., 2007). As the students primarily get feedback from the materials and their peers, teachers must relinquish control and allow students to work independently for certain periods of time. When acting as a partner, "teachers can mediate their pedagogical expertise without shutting down ongoing activities"

(Liljeström et al., 2013, p. 79). However, the literature is equivocal with regard to the characteristics of the teacher's role in open-ended S&T learning environments (Hakkarainen, 2009; Kolodner, 2001). Several teacher aspects are considered important for this role. Firstly, the asking of authentic questions which have no pre-specified answers; as well as higher-level questions, and student-generated questions in particular, as they are substantively engaging (Nystrand, Wu, Gamoran, Zeiser, & Long, 2003). Secondly, a teacher needs to be able to integrate students' everyday experiences and raise these experiences as resources that can be shared by the whole learning community (Viilo et al., 2011). Thirdly, classroom management (Jacobowitz, 1997; Lawson, 1995) and an efficient use of classroom time remain important in these open-ended settings. Finally, a few authors (e.g. Doppelt, 2003) point to the importance of a positive climate in project-based learning environments in S&T. However, different studies in other fields of education have already shown that when teachers create a sense of community, respond to students' needs and foster positive relationships, students are more likely to be engaged in the learning process (Marks, 2000; Rimm-Kaufman, La Paro, Downer, & Pianta, 2005).

We will investigate the role of teacher-pupil interactions as well as the interactions among pupils. In order to that, we will summarise how previous studies conceptualised these interactions. A number of recent studies analyse classroom discourse (Chin, 2007; Erdogan & Campbell, 2008; Hackling et al., 2011; Reinsvold & Cochran, 2012; Scott et al., 2006; Smart & Marshall, 2013) when studying teacher-student and pupil-pupil interactions. In discourse analysis various methods are used which focus on the written and spoken language of these interactions (Mercer, 2010). The nature of the questions asked in the classroom – both by teachers and students – is studied, and the teacher's elaboration of students' answers is evaluated. As such, one can study whether questions elicit and scaffold students' ideas (Smith et al., 1993) and to what extent the teacher keeps a leading role in classroom discussions. While these studies have their value for the investigation of interactions, we argue that it is interesting to take a broader perspective by studying also the emotional, organisational and instructional aspects of interactions that are part of or closely relate to this classroom discourse. Both the verbal and nonverbal interactions between teachers and students can be taken into account, and provide insights into the quality of the relations in class, whether teachers and students build on (other) students' responses, a teacher's sensitivity, a teacher's openness towards students' ideas and responsibilities, his/her management of student behaviour and classroom activities, and so on (Pianta et al., 2012).

Teachers' attitudes towards S&T and its teaching

It is not only the interactions between teachers and pupils and among pupils which can play a role in pupils' engagement in open-ended S&T learning environments. Teachers' underlying attitudes towards S&T (and its teaching) may also affect pupils' engagement.

An attitude is "a learned disposition to respond in a consistently favourable or unfavourable manner with respect to a given object" (Fishbein & Ajzen, 1975, in Young, 1998, p. 97). The concept of attitude towards science is often poorly articulated (Barmby, Kind, & Jones, 2008; Bennett, Rollnick, Green, & White, 2001; Coulson, 1992; Osborne et al., 2003; Pajares, 1992). However, in the literature some recurring distinctions have been made.

Firstly, a distinction can be made between a teacher's personal attitude towards S&T and his/her professional attitude with regard to the teaching in these domains (Oberon, 2009; Van Aalderen-Smeets et al., 2011; Wilkins, 2008). Asma and colleagues (2011) found out that teachers perceive a distinction between their personal and professional attitude toward science and that it is possible for these attitudes to develop independently. While the personal attitude encompasses the attitude of the teacher as a citizen, independent of their profession, the professional attitude involves the attitude towards the teaching of S&T[21] in the context of the school (Van Aalderen-Smeets et al., 2011). This last aspect deals with where the teacher stands on how S&T is presumed to be taught nowadays (for example, by letting pupils experiment, solve problems on their own, design their own products…).

Secondly, and similarly to the concept of engagement, an attitude is often divided into three dimensions: affective, cognitive, and behavioural (Eagly & Chaiken, 1993; Katz & Stotland, 1959; Klop & Severiens, 2007; Rosenberg & Hovland, 1960). Van Aalderen-Smeets et al. (2011) demonstrated the existence of different attributes for each of these three dimensions. The dimensions are quite similar for both the personal and the professional attitude of teachers towards S&T.

The affective dimension of a teacher's attitude towards S&T consists of feelings and moods that he/she experiences in relation to these domains (Van Aalderen-Smeets et al., 2011). As often conceived in the literature, this dimension deals with how joyful a teacher feels about S&T (Palmer, 2004; Young, 1998) and with the general enjoyment of teaching science (Johnston & Ahtee, 2006; Ramey-Gassert, Schroyer, & Staver, 1996).

[21] While the teaching of science is often called 'inquiry' teaching, the teaching of technology is often called 'design' teaching.

The cognitive dimension consists of evaluative thoughts and beliefs towards science, technology, inquiry and design. This dimension firstly contains the perceived relevance or importance of science, which refers to "the extent to which people consider science relevant or important for their personal lives, for society, for prosperity, or for health" (Van Aalderen-Smeets et al., 2011, p.164) as well as the perceived relevance of teaching science (Appleton & Kindt, 1999; Carleton, Fitch, & Krockover, 2008; Cobern & Loving, 2002; Johnston & Ahtee, 2006; Liang & Gabel, 2005). Secondly, the cognitive dimension contains the perceived difficulty of S&T, which refers to thoughts and beliefs concerning the general difficulty of science relative to other fields of study, and their perceptions about the difficulty of teaching science (Harlen & Holroyd, 1997; Johnston & Ahtee, 2006; Liang & Gabel, 2005)[22].

The behavioural dimension constitutes the behavioural responses or actions of a person when confronted with science, technology, inquiry and design. This response can be either overt (with the person actually acting out the behavioural response or action) or covert (with the person intending to act out the behavior, although the action has yet to take place). While this dimension is seen as part of the attitude concept, Van Aalderen-Smeets et al. (2011) found no articles concerning primary school teachers' attitudes that reported having measured behaviour related to science in respondents' daily lives or behavioural intention to engage in activities related to S&T. However, according to the authors, a few studies investigated the behavioural component of attitude towards teaching science (Appleton & Kindt, 1999; Goodrum, Hackling & Rennie, 2001; Haney, Czerniak, & Lumpe, 1996; Palmer, 2001; Yates & Goodrum, 1990).

In their recent review, Van Aalderen-Smeets et al. (2011) proposed a new theoretical framework for primary school teachers' attitudes towards science. In this framework they appeal to the Theory of Planned Behavior (Ajzen & Fishbein, 1980), in which behavioural intention is viewed as a direct outcome of the cognitive and affective dimension of attitudes, and not as a component of attitude itself[23].

[22] Besides these different aspects Van Aalderen-Smeets et al. (2011) also distinguish thoughts and beliefs with regard to gender roles in S&T (e.g. the perception that men are better at understanding S&T than women).

[23] In the new theoretical framework proposed by Van Aalderen-Smeets et al. (2011), another affective dimension – that of Anxiety – is also distinguished. Furthermore, the 'Perceived Control' dimension – consisting of Self-Efficacy and Content Dependency – was added.

The relation between interactions, teachers' attitudes towards S&T (and its teaching) and pupils' engagement (growth)

Several studies have shown the effects of open-ended learning environments on pupils. These learning environments have shown to be beneficial for students' engagement (Kaldi et al., 2011; Mant et al., 2007; Wurdinger, Haar, Hugg, & Bezon, 2007). The majority of these studies focused on the effectiveness of the typical characteristics of these learning environments. Cornell and Clarke (1999), for example, found that students were more engaged when involved in project-based learning environments than in teacher-directed ones; the students reported that this is due to being given the opportunity to work with other pupils while doing hands-on activities. In a study by Mant et al. (2007), teachers followed a continuing professional development programme in which they were trained to implement cognitively challenging, practical science lessons with plenty of space for thinking and discussion. By conducting focus group interviews, the authors found that 10- and 11-year old pupils had a higher level of enthusiasm and engagement, which they attributed to more experiments and investigations, new discussion activities, more thinking for themselves and less time spent on writing. According to the respondents, their engagement in learning was increased by the encouragement to think and the challenges provided by the teachers. They enjoyed doing active things, discussing their ideas with each other (particularly when working in small groups), and they appreciated that the teachers were encouraging them to think more for themselves; in particular, when they were encouraged to use their own ideas in investigations, hence increasing their autonomy. Wallace (1996) and Osborne and Collins (2000) also concluded that giving room for autonomy is important for pupils' engagement.

This study adds to previous research in different ways. Firstly, in the above studies, engagement is mostly conceptualised as affective engagement, without integrating the behavioural and cognitive dimensions into the concept. Secondly, an observation scale for the assessment of pupils' engagement has, to our knowledge, never been used before when investigating the effects of a project-based S&T learning environment. In the study by Mant et al. (2007), for example, focus group interviews were conducted in order to evaluate pupils' own perceptions of their evolution in engagement instead. Finally, research in the middle school context gives indications that the core characteristics of S&T learning environments – i.e. students' autonomy and collaboration among students – do not guarantee successful learning in these learning environments. For instance, in a study by Scardamalia and Bereiter (1992), students tended to generate low-level factual questions rather than questions that could extend their understanding of a topic, scarcely commented on each other's questions, and gave superficial rather than constructive feedback. Various other authors have found that students do not discuss ideas or use evidence systematically

(Palinscar, Anderson, & David, 1993; Germann & Aram, 1996). While pupils may be affectively engaged, they are not necessarily cognitively engaged during challenging S&T activities. Researchers such as Keys and Bryan (2001) and Krajcik and colleagues (1998) have called for an investigation into how teachers can make these environments successful for students' engagement and learning via their actions in S&T learning environments. In this study we will try to answer this question.

It is plausible to expect that not only interactions can play a role in pupils' engagement. The underlying and less visible attitudes of teachers towards S&T (teaching) may also affect pupils' engagement. Previously, different studies have shown the effects of teachers' attitudes towards S&T (teaching) on pupils' attitudes towards the domains of S&T (Harlen & Holroyd, 1997; Jarvis & Pell, 2005), but not in relation to (the growth in) pupils' engagement during S&T activities. Pupils' levels of engagement are closely related to their attitudes, and indeed can be perceived as a concrete manifestation of these attitudes.

Research context

The data for the present study was collected in the context of an intervention study in which the project-based S&T learning environment 'Village@School' was implemented. Village@School is a project which was designed and implemented in 2008, born out of a collaboration between the Centre for Experiential Education (De Winter et al., 2010), IMEC and the Roger Van Overstraeten (RVO) society. In Village@School[24], pupils are challenged to build a miniature village – or any other site, such as a theme park, for example - with as many (working) applications (e.g. working traffic lights) as possible, in a minimum of 10 weeks (20 lessons, each of 2 hours). In principle, pupils are permitted a lot of room for initiative in this project, as they are responsible for the decisions about what will be built on the plate and how this will be done. With a budget of one hundred euros the class can buy whatever materials they consider necessary. It is encouraged that they use recyclable materials, as well as the so-called 'physical tools' (e.g. Liljeström et al., 2013) – such as electric drills and hammers – and 'cognitive tools' for finding information (e.g. books, maps, the Internet). At different points during the project support was given to the teachers. This support was expected to be necessary as Flemish and Dutch teachers often have no specific affinity with S&T education. The project was launched in an introductory conference and workshop. While conducting the project with their pupils, the teachers were given a second workshop and two coaching sessions. In this second workshop, teachers reflected on their experiences with their colleagues and the research team. The coaching

[24] More information with regard to this project as it was originally implemented can be found on this website: http://www.dorpopschool.be. In this study, the project principles and research activities are slightly different. More updated information about the project and this study can be found in De Winter, Thys, & Stas (2013).

sessions were designed to meet the teachers' needs and aimed at supporting them in their guiding role during the project. During each of these sessions teachers were encouraged to give pupils autonomy and make stimulating interventions, such as posing thought-provoking questions while participating in project activities.

In light of our theoretical framework and the sketch of the research context, our first research question is: Is there a growth in pupils' engagement in the course of the implementation of the project Village@School? (research question 1). Secondly, can we explain the possible differences between schools/classes concerning their growth in pupils' engagement by means of teachers' attitudes towards S&T (teaching) and the quality of interactions? (research question 2).

Method

Participants

Initially, 34 primary school classes within 18 schools participated in this study. Four of these schools were located in the Netherlands (8 Dutch classes), and 14 in Belgium (26 Belgian classes). The participating teachers taught in the 3rd grade (3 groups of pupils), 4th grade (7 groups of pupils), 5th grade (16 groups of pupils), 6th grade (16 groups of pupils), or in a combination of two consecutive grades. Four classes contained both 5th and 6th grade pupils; three classes had both 4th and 5th grade pupils and three others consisted of both a 3rd and 4th grade pupils. Two of these mixed classes - a 3rd and 4th grade and a 5th and 6th grade - belong to a school for highly gifted pupils. Shortly after the pre-measurement and before the start of the Village School project, some teachers could no longer participate or chose not to participate in the study (five in total). We decided to let two classes and their 'replacing' teachers take part in the further measurements of the study[25]. This resulted in a new sample of 31 teachers and their classes. Two teachers from the same school took part in the study; except for two cases in which only one teacher from the school participated and one case in which three teachers of one school were engaged (one of these teachers dropped out). In the course of the study, another teacher dropped out. Including those who dropped out, a total of 36 teachers participated in this study; 26 of whom were female, and 10 were male. In total, 613 pupils took part. However, data was collected on a sample of ten randomly selected pupils per class: five boys and five girls. Pupils who were absent for either of the measurements were replaced by randomly chosen peers.

For practical reasons, data was collected in two waves. The first part of the data collection (involving 19 classes) started in November 2013; the second part (involving 17 classes) in September 2014. During the pre-measurement, teachers' initial attitudes, the quality of the interactions, and pupils' engagement were evaluated. During the intervention, the quality of the interactions and pupils' engagement were measured. In the post-measurement, the quality of interactions and pupils' engagement were evaluated, along with teachers' final attitudes.

[25] The two substituting teachers who joined the study at a later time, couldn't follow the introductory conference anymore but were sent the recorded video clip. Also the two teachers who – because of other commitments - could not participate to the second workshop, were sent the recorded video clip. Although the researchers urged the teachers to watch these video clips, we cannot guarantee they effectively watched these videos.

Data Collection

The data was collected by means of classroom observations and a teacher questionnaire. In what follows, these instruments will be presented and results with regard to their validity and reliability will be reported.

Classroom Observations

In order to answer research question 1 and 2, both before and after the intervention primary school teachers conducted two equivalent, standardised S&T assignments (called the Bridge activity and the Tower activity) with their pupils (Smeets, 2014). Before the Village@School project, teachers conducted the activity 'Building a bridge' and after the project they implemented the 'Building a tower' activity with their pupils. During these activities pupils' engagement and the quality of the interactions were evaluated (research question 2). Teachers were given approximately one week to prepare themselves to implement this activity in their classroom. The Bridge and Tower activities were assignments on which a maximum of 45 minutes could be spent, and in which a bridge and tower respectively needed to be constructed. These equivalent activities had to fulfil the following requirements: only paper strips (width: 5.25 cm), staplers, adhesive tape and glue sticks could be used. Moreover, the bridge had to cover a distance of thirty centimetres, and the strength of the construction had to be tested by placing a small, plastic bottle, filled with two hundred and fifty millilitres of water, on top of it. Likewise, the tower had to be made as high as possible and also had to support a full bottle with the same volume of water. Except for these requirements and the time restriction (45 minutes), the teachers could choose the way in which they organised and realised the activity with their pupils.

During each of these activities and during two sessions of the implementation of the Village@School project, in which teachers and pupils worked on the project[26], pupils' engagement was coded live by an observer via the Leuven Involvement Scale - Primary (LIS-P) (Laevers, 2011). Applying the 'scanning procedure', the engagement of the ten randomly selected pupils were scored subsequently in two cycles while classes conducted the activities 'Building a bridge', 'Building a tower' and during two one-hour Village@School sessions. After each two-minute observation period, pupils' engagement was rated on a scale from 1 (no involvement) to 5 (non-stop, very intense involvement) (Doumen, Koomen, Buyse, Wouters, & Verschueren, 2012). Half-scores could also be given. Scores across observation periods were averaged to obtain an overall observer-rated indicator of children's engagement for the standardised

[26] The first observation took place shortly after the start of the Village@School project; the second observation took place after the second workshop and first coaching session (the latter was given by a colleague).

activities, or in the case of Village@School, for one session during the project. In previous studies, the LIS proved to have excellent inter-rater reliability (r = .75 to .90) in nursery to primary school settings (Laevers & Laurijssen, 2001; Van Heddegem, Gadeyne, Vandenberghe, Laevers, & Van Damme, 2004).

In the current study, the single-measure intraclass correlation was calculated for four measurement rounds to assess the amount of agreement: firstly between the scores from Observer 1 and from Observer 2, and secondly between the scores given by Observer 1 at a time 1 and at a time 2. For the pre-measurement of the first wave the inter-rater reliability of Observer 1 with Observer 2, each rating involvement of ten children[27], was excellent (ICC = .79; p = .000)[28]. For the pre-measurement of the second wave, for the two Village@School observations of the second wave and for the post-measurement of the second wave, intra-rater agreement was calculated for 35 pupils (Observer 1). ICCs also proved excellent, ranging from .93 to .97.

While pupils' engagement was scanned by Observer 1, the quality of the interactions was coded by another observer (Observer 3) using the Classroom Assessment Scoring System Upper Elementary (CLASS) (Hamre & Pianta, 2007). Training is required to become a reliable CLASS observer (http://teachstone.com/services/training/class-observation-training-programs). Observer 3 followed intensive CLASS training, obtained a certificate and passed an annual renewal test. Evidence suggests that CLASS scores, when assigned by trained, certified observers, are highly reliable (Pianta et al., 2012).

The starting point in CLASS is the interactions between the teacher and the pupils, but the interactions among pupils are also rated (Luckner & Pianta, 2011). Via their interactions with pupils, "teachers act as invisible hands in the classroom, influencing children's behavior with their peers by modeling relational skills, organizing and facilitating opportunities for peer interactions in the classroom, as well as teaching skills that indirectly relate to peer behavior (e.g. regulatory skills and language and cognitive skills)" (Luckner & Pianta, 2011, p.257).

According to the CLASS, interactions in the classroom occur in three different domains: emotional support, classroom organisation, and instructional support (Hamre & Pianta, 2007; Hamre, Pianta, Mashburn, & Downer, 2007; Pianta, La Paro, & Hamre, 2008), which are further subdivided into dimensions (see Table 1). Large-scale studies provide evidence for the existence of these three different domains of interactions (Hamre et al., 2007). Research shows associations of these domains with children's

[27] The observations were re-scored on the video recordings to calculate inter/intra-rater reliability, which we initially made in order to score using the CLASS.
[28] One observer coded the observation live, the other one on video. The live observer also recorded the engagement of the randomly selected pupils for about 2 minutes.

achievement (e.g., Pianta et al., 2008) and with their social/emotional and behavioural functioning (e.g. Howes, 2000; NICHD ECCRN, 2003).

Firstly, the emotional support domain encompasses interactions that reflect the emotional climate of the classroom, called 'Positive Climate', which is conceptualised as the warmth and/or negativity present in the interactions, as well as the emotional connection between the teacher and the pupils. Emotional support also includes 'Teacher Sensitivity', or a teacher's awareness of and responsiveness towards pupils' levels of academic and social/emotional functioning and their developmental needs (NICHD ECCRN, 2002). Next, the extent to which a teacher shows openness towards pupils' ideas and opinions and the extent to which he/she provides opportunities for responsibility belong to this domain ('Regard for Student perspectives'). Secondly, the classroom organisation domain includes the teacher-pupil interactions involved in managing time, behaviour, and attention in the classroom (Hamre & Pianta, 2007; Pianta et al., 2008). These interactions include teachers' attempts to effectively manage class time ('Productivity'), prevent and redirect misbehaviour (including misbehaviour directed to peers) ('Behaviour Management'), and direct pupils' attention through clear and consistent organisational systems. Negative interactions characterised by sarcasm, frustration and harsh voices have to be limited ('Negative Climate'). Thirdly, the instructional support domain contains the quality of the instructional interactions between teachers and pupils in terms of the richness of the instruction and feedback provided (Hamre & Pianta, 2005, 2007; Pianta et al., 2008). In this domain it is measured how well information is presented, e.g. by using a variety of modes and materials ('Instructional Learning Formats'); how many interactions promote a deep level of understanding in pupils, e.g. by making real world connections ('Content Understanding'); to which extent pupils get opportunities to experiment, brainstorm and reason using higher-order thinking skills without direct guiding from the teacher ('Analysis and Inquiry'); how many (back-and-forth) exchanges occur that scaffold and encourage pupils' thinking and learning ('Quality of Feedback'); and finally whether profound dialogues among pupils are stimulated ('Instructional Dialogue').

The dimensions, on their turn, are subdivided into indicators and behavioural markers. The instructional support domain, for example, consists of the 'instructional dialogue' dimension, subdivided into three indicators: 'cumulative content-driven exchanges', 'distributed talk' and 'facilitation strategies'. The indicator 'distributed talk' consists of four behavioural markers: student-initiated dialogues, balance of teacher and student talk, and a majority of student and peer dialogues. Using the indicators and markers, for each dimension a score is given on a 7-point scale. The indicators are situated in the low, mid or high range, and the behavioural markers help to situate these indicators in one of these ranges.

Table 1. The CLASS Domains (Emotional Support, Classroom Organisation and Instructional Support) with their Dimensions.

Emotional Support	**Classroom Organisation**	**Instructional Support**
Positive Climate	Behaviour Management	Instructional Learning Formats
Teacher Sensitivity	Productivity	Content Understanding
Regard for Student Perspectives	Negative Climate	Analysis and Inquiry
		Quality of Feedback
		Instructional Dialogue

Note: More detailed descriptions of the domains and their dimensions can be found in Table 1 of Hafen, Bridget, Hamre et al. (2015)

All classroom observations were video-recorded. Lessons were recorded with one camera with a wireless microphone attached to the teachers. This made it possible to capture teachers' interactions with pupils precisely, which is necessary in group-work settings in which pupils are talking with each other and interactions with the teacher are not always clearly audible. As well as the interactions with the teacher, interactions among pupils were also recorded. The camera was positioned sideways in the classroom, in order to have a global view of the class. Due to time constraints, the quality of the interactions in the pre- and post-measurement was coded on the basis of the video material[29]. The Village@School sessions were coded live[30].

With the CLASS, two observation cycles were conducted in the activities 'Building a bridge' and 'Building a tower'. Each observation cycle consisted of 15 minutes' observation followed by 10 minutes' scoring. To obtain one general score for each teacher for each CLASS dimension, the available scores assigned in each observation cycle were averaged. Domain scores were calculated by averaging the dimension scores. Two Village@School sessions (approximately a 1-hour observation) were coded live, which resulted in two observation cycles for each session. However, for reasons of practicality during the project (e.g. teachers stopped the session early, etc.) some cycles could not be fully conducted because there

[29] The observations of the post-measurement of the first wave were coded live (except for the dimension Negative Climate which was coded on the video).

[30] Two other observations were video-recorded by a colleague, after which a coaching session was provided (one after the second workshop, and one last observation near the end of the project). These observations were not taken into account for this particular study.

was not enough time left after scoring the first 15 minutes of the observation. As the video recording allows you to stop watching the video after a 15-minute observation and start again at that time after scoring the first observation, it was possible to score more cycles via the video or to select another cycle (not containing the routine activities such as tidying up). It is important to note that the video scores did not differ significantly from the live scores. A paired samples t-test showed no significant p-values for the three CLASS domains ($p = .427$ for emotional support, $p = .457$ for classroom organisation and $p = .694$ for instructional support) or for most of the dimensions (except for Regard for Student Perspectives, Productivity and Instructional Dialogue). The paired samples t-test however is a rather strict test, which does not take into account the margin one has when scoring with CLASS. CLASS has been used before to code videotapes of classrooms (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Kane & Staiger, 2012). When it was exceptionally still not possible to score two cycles, for example because teachers conducted the activity 'Building a bridge' in less than 45 minutes, a missing score was given for this session. This was done because more cycles are needed in order to draw reliable conclusions about the quality of the interactions (Pianta et al., 2012). The pupil engagement data for which we had one or more missing predictor variables were excluded from the dataset in the further analyses.

To assess intra-rater reliability[31], CLASS dimensions were firstly evaluated again for 38 lessons over all 4 measurement rounds (the two S&T assignments, and the two measurements during Village@School). Following the criteria from Cichetti and Sparrow (1981), the ICC for the three domains varied from good to excellent (.61 for emotional support, $p = 0.000$, .84 for classroom organisation, $p = 0.000$; and .68 for instructional support, $p = 0.000$). For the dimensions the intra-rater reliability was fair to excellent, as ICCs ranged from .42 to .90; except for the dimension 'instructional learning formats', which had an ICC of .26[32]. In other studies using CLASS, ICCs (two observers) ranging from.15 to .43 are reported (Hafen et al., 2015); these lower ICCs are partly due to the fact that in large scale studies with CLASS, adjacent scoring was allowed (Pianta et al., 2012).

Secondly, CLASS dimensions were evaluated again for the pre-measurement, both of

the Village@School observations, and the post-measurement separately. For the pre-measurement (18 teachers), the ICC for the three domains varied from fair to excellent (.51 for emotional support, $p = .046$; .92 for classroom organisation, $p = .000$; .78 for instructional support, $p = .001$). The ICC for the domains in the Village@School observations was fair to good (ICC = .47 for emotional support, $p = .024$; ICC = .64 for classroom organisation, $p = .009$; ICC = .66 for instructional support, $p = .003$) and in the post-measurement (10 cases) the ICC varied from fair to excellent (ICC = .44 for emotional support, p =

[31] Because of the high cost of having the data double-scored by another trained and certified researcher, it was not possible to calculate inter-rater reliability. However, reliability is assured by following training, being certified and recertified.

[32] Therefore, analyses with this dimension were excluded.

.088; ICC = .92, p = .000 for classroom organisation and ICC = .67, p = .013 for instructional support). As intra-rater reliability for different dimensions across the separate measurements (pre-measurement, Village@School sessions and post-measurement) was poor according to the criteria of Cichetti and Sparrow (1981), no analyses with the dimensions in the separate measurement rounds were conducted.

Teacher Attitude Questionnaire

To answer research question 2 an attitude questionnaire developed by Oberon (2011) was administered. Via this questionnaire teachers' attitudes towards S&T and their teaching in the fields (as part of the pre-measurement) were measured. In line with the literature, the cognitive, affective and behavioural dimensions of attitude towards S&T (teaching) are represented in the questionnaire. The questionnaire consists largely of a sample of items from the VTB-monitor as developed by Walma van der Molen (2007). It contains two subscales for teachers' personal attitudes (labelled 'attitude towards technology' and 'attitude towards science') and for their professional attitudes (labelled 'attitude towards design learning' and 'attitude towards inquiry learning'), of which the items had to be rated on a 5-point Likert scale ('completely disagree', 'don't agree', 'neutral', 'agree', 'totally agree'). The theoretically distinguished dimensions in the attitude concept are divided into these scales. Firstly, the affective dimension is evaluated via the 'pleasure' scale (e.g. 'I find technology interesting'; 'I prefer to leave children to their own devices, rather than tell them precisely how something works'). Secondly, the cognitive dimension is operationalised in the 'difficulty' scale (e.g. 'I find doing inquiry hard'; 'For a teacher it is hard to find an appropriate problem that may be the starting point for design learning') and the 'importance' scale (e.g. 'Technology is important for society'; 'Already in primary school children have to think in an inquiry-based way'). Finally, the behavioural dimension is represented via the 'intention future' scale (e.g. 'I like reading about new inventions, for example in the newspaper or on the Internet'; 'I would like to try out activities concerning design learning in class'). Next to these attitudes towards S&T (teaching), teachers' self-efficacy is measured in the questionnaire, but was left out in the analyses. After conducting factor analyses and examining Cronbach's alpha, Oberon (2011) left out some scales (and items), i.e. the difficulty scales of the attitudes towards design and inquiry learning, the intention future scale of the attitude towards design learning, and the importance scale of the attitude towards science. In the end, Cronbach's alpha for the remaining scales of the four different attitudes (science, technology, design learning, inquiry learning) varied from .60 to .87. Despite the fact that Oberon (2011) left out some scales, we decided to retain the initial questionnaire scales in our analyses, in line with the theoretical tripartite division of the attitudes in dimensions (Eagly & Chaiken, 1993; Katz & Stotland, 1959; Klop & Severiens, 2007; Rosenberg & Hovland, 1960), and conduct factor analyses on our own data (see further 'data analysis'). The items that initially belonged to the self-efficacy scale, but loaded high on the difficulty scale in the study of Oberon (2011) were also retained in the analyses. When an item that theoretically belonged to the difficulty scale, but loaded high on the self-efficacy scale and not on the difficulty scale (Oberon, 2011), this item was removed from the analyses.

Data analysis

Factor analyses were conducted to determine the structure of the attitude questionnaire. Exploratory and confirmatory factor analyses were performed for the personal attitudes (one for science and one for technology) and for the professional attitudes (one for design learning and one for inquiry learning). As factor extraction method we chose the Principal Component Analysis in SPSS. In order to be able to interpret the data structure, we made use of oblique rotations. When on the basis of the output a choice had to be made about the number of factors, not only the Kaisercriterium (eigenvalue > 1) was taken into account. The second criterium was the result obtained via the scree test. Based on the number of factors indicated by the scree plot in het output of the exploratory factor analyses, confirmatory factor analyses were conducted. Items loading below .30 were considered 'low' as low factor loadings.

Given the nested structure of the data (pupils' engagement engagement is situated at the pupil level, but may also be determined by the class and school level), the data were analysed by means of multilevel modelling techniques (Goldstein, 1995), making use of the software program MLwiN (Rashbash, Charlton, Browne, Healy, & Cameron, 2005). As we had data available for four subsequent moments on all variables apart from teachers' attitudes, which were only measured in the pre-measurement, time can be considered as an additional predictor at an additional level in the multilevel models (Singer & Willett, 2003). Four levels of information are involved in this study: time (level 1), nested within pupils (level 2), nested within classes (level 3) and nested within schools (level 4).

Firstly, baseline models were tested in the prediction of pupils' engagement to partition the variance for each of the outcomes between school, class, pupil, and time level. This was done for all four moments of measurement together (pre-measurement, two measurements during implementation of Village@school and post-measurement). Where there was a significant level of variance at the time level, a predictor time - centred around the grand mean - was created and tested in the model. In case of a significant time predictor, we allowed the effect of time to vary across classes as we were interested in the possible differential effect of classroom variables across time. When significant, the differential effect of teachers' attitudes towards S&T (teaching) on pupils' engagement across the subsequent measurement rounds was determined (Models A). This was not done for the varying CLASS domains and dimensions as only stable characteristics could be varied with time.

Secondly, after conducting a paired samples t-test, a new variable – pupils' growth in engagement, being the difference between pupils' engagement in the post- and pre-measurement (post scores – pre-scores) – could be created. Again, a baseline model could be tested in the prediction of pupils' growth in engagement. In case of a growth in pupils' engagement, the scores for teachers for each CLASS domain/dimension for each moment of measurement could be averaged[33] and added to Models C and D as predictors of pupils' engagement growth (post-pre). Also for the new outcome variable, each attitude was added in Models B.

Due to additional missing data on some predictor variables, the sample size is sometimes smaller throughout the analyses. The exact number of cases on which results are based is reported for every analysis separately.

[33] This could not be done for the whole sample of teachers, as scores for all teachers were not available for every measurement occasion. For 23 teachers an average score over all measurement rounds could be calculated.

Results

Factor analyses- attitude questionnaire

Both exploratory and confirmatory factor analyses did not reveal clear evidence for the hypothesised factor structure. In a next step, we decided to leave out the behavioural items of both teachers' personal and professional attitudes. This is justified, because as discussed in our introduction, there is yet no consensus that actual behaviour or the intention to conduct that behaviour is part of the attitude concept (Van Aalderen-Smeets et al., 2011). After omitting the behavioural items, an exploratory factor analysis for the Attitude towards Technology resulted in a scree plot with two factors. Next, a confirmatory factor analysis was conducted. As items of two of the three retained scales - the 'Importance' scale and the 'Difficulty' scale - theoretically belong to the cognitive component of attitudes, it was plausible to expect their corresponding factors in the confirmatory analysis. However, a mix was found of items of the 'Pleasure' scale and the 'Difficulty' scale on the one hand, and of the 'Importance' scale on the other hand. Results pointed in the direction of one, single underlying factor, containing cognitive and affective elements for the respective attitudes towards science, technology, inquiry learning and design learning, containing affective as well as cognitive elements. This is in line with the literature, and more specifically with the Theory of Planned Behavior (TPB) of Ajzen and Fishbein (1980) according to which attitude only consists of a cognitive component and an affective component. In a next step confirmatory factor analyses, set on one factor, were conducted. This was also done for the Attitude towards Design Learning, the Attitude towards Science and the Attitude towards Inquiry Learning after conducting exploratory factor analyses which did also not result in the theoretically expected factors. The factor for the Attitude towards Technology explains about 31% of the total variance between the scores on the items; for the Attitude towards Design Learning the factor explains about 26% of the variance; the retrieved factor for the Attitude towards Science explains about 27% of the variance and the one for the Attitude towards Inquiry Learning explains 26% of the variance. The different items with their factor loadings on the four attitudes are presented in Table 2.

Table 2. The Items of the Attitude Questionnaire with their Factor Loadings on the Attitudes for Technology (T), Design Learning (DL), Science (S) and Inquiry Learning (IL)

	Attitude T	Attitude DL	Attitude S	Attitude IL
I find technology interesting (P)[a]	.715[b]			
Technology is important for society (I)	.058			
The Belgian government has to spend More money on technology (I)	.268			
I find it onerous to repair something by myself (P)	.837			

	Attitude T	Attitude DL	Attitude S	Attitude IL
Technology has a great influence on people (I)	-.049			
I like designing things on my own (P)	.724			
I like putting things together (P)	.773			
Technology has a negative influence on the society (I)	-.273			
The subject of technology can only be taught by special educated teachers (D)	.662			
I like repairing things myself (P)	.918			
When a country does a lot about technology, this is good for the economy (I)	.091			
Of myself I have no interest in technology (P)	.800			
Technology makes our life nicer (I)	-.189			
Technology is good for the economy of Belgium (I)	-.181			
In order to teach technology properly, you first have to attend a specialized training (D)	.451			
It requires a lot from a teacher to let design learning go quite smoothly (D)		.671		
It is important that yet in primary school children learn how to solve technological problems (on their level) (I)		.154		
I like letting children solve technological problems (P)		.666		
It is better to leave design learning with children to an experienced teacher (D)		.585		
It is difficult to apply design learning i in primary school (D)		.749		
It is difficult for a teacher to find a suitable problem that may be the starting point for design learning (D)		.048		
It is important for society that children learn to think of technological solutions for a specific problem (I)		-.201		
Design learning in primary school is necessary to prepare children for secondary education (I)		.399		
I find it interesting to let children solve a technological problem via design learning (P)		.586		
I like inventing things (P)			.419	

	Attitude T	Attitude DL	Attitude S	Attitude IL
A good teacher needs to have an inquiry attitude (P, I)			.071	
I like to think of new ideas (P)			-.050	
I find scientific research interesting (P)			.692	
To do research, you have to be decently smart (D)			.262	
I like clearing up things (P)			.619	
Researchers do important work (I)			.340	
I find doing inquiry hard (D)			.857	
I find it important to know how things are put together (I)			.338	
Science is complicated (D)			.721	
It is a great effort for a teacher to have children learn inquiry in class (D)				-.103
It is important that children from a young age on learn how they have to do inquiry (on their level) (I)				.161
I like it more to leave children to their own devices, rather than that I tell them precisely how something works (P)				.508
I find it interesting to let children Invent things themselves without using explanations from a schoolbook (P)				.777
I find it hard to apply inquiry learning primary school (D)				.226
I like letting children clear things up (P)				.805
Yet in primary school children have to think in an inquiry way (I)				.468
You only have to start with inquiry learning with children when you are an experienced teacher (D)				.185
I like letting children do new discoveries by themselves (P)				.708
It is important for the society that children learn to think in an inquiry way (I)				.700
I doubt whether I have enough skills to have children learn inquiry (D)				.301
Inquiry learning in primary school is necessary to prepare children properly secondary school (I)				.133
I like children letting invent things by themselves in class				.686

Note: The items that were negatively formulated were reversed in the analyses. For some items, low and/or negative loadings were obtained. In terms of consistency, we wanted the cognitive and affective component of each of the four attitudes to be equally represented in the four attitudes. It was decided not to leave out these items after conducting the factor analyses. After calculating the internal consistency, the items in bold were used to calculate the mean for each of the attitudes.

[a] The characters between the brackets refer to the subscales to which the items theoretically belong. The pleasure scale = (P); the importance scale = (I); the difficulty scale = (D).

[b] The column in which an item's loading is presented refers to the Attitude (T, DL, S, IL) to which the item belongs.

The internal consistency for each of the four attitudes ('attitude towards technology', 'attitude towards science', 'attitude towards inquiry learning', 'attitude towards design learning') was calculated. Cronbach's Alpha was .76 for the attitude towards technology, .69 for the attitude towards science, .66 for the attitude towards design learning, and .69 for the attitude towards inquiry learning. The minimum level of alpha of a good test is normally .70 (Gable & Wolf, 1993; Kline, 2000; Nunnally, 1978). However, an alpha starting from .60 can be considered acceptable. A composite score was formed by averaging the scores on the items for each of the four attitudes (the items in bold in Table 2).

Correlations between the four attitudes can be found in Table **3.** Oberon (2011) also found that the higher teachers' personal attitude towards technology, the higher their personal attitude towards science. The same is the case for the attitudes towards design and inquiry learning.

Table 3. Correlations between the Four Attitudes

	Technology	Science	Design Learning	Inquiry Learning
Technology	1	.65**	.66**	.21
Science		1	.53**	.19
Design Learning			1	.36*
Inquiry Learning				1

Note: $^{*}p \leq 0.05$, $^{**}p \leq 0.01$; $^{***}p \leq 0.001$

Relation between teachers' attitudes towards S&T (teaching), time and pupil's engagement (all measurement occasions)

Table 4 presents descriptive statistics and correlations for the first outcome variable (Pupil's Engagement).

Table 4. Means, Standard Deviations, Range, Skewness, Kurtosis and Bivariate Correlations Between Study Variables (Pupils' Engagement as outcome)

	M	SD	Range	Skewness (SE)	Kurtosis (SE)	1	2	3	4	5	6	7	8	9
Engagement_all_measurement_occasions	3.26	.73	4.00	-.56 (.08)	.28 (.15)	1								
Engagement_measurement_occasion_1	3.16	.77	3.75	-.72 (.15)	.28 (.29)		1							
Engagement_measurement_occasion_2	3.19	.70	4.00	-.10 (.15)	-.06 (.29)		.18**	1						
Engagement_measurement_occasion_3	3.26	.73	3.75	-.63 (.15)	.30 (.29)		.13*	.19**	1					
Engagement_measurement_occasion_4	3.45	.67	3.75	-.71 (.15)	.84 (.31)		.19**	.23**	.18**	1				
Attitude_T	3.80	.41	1.93	-.72 (.42)	1.06 (.82)	.02	.10	.03	.02	-.07	1			
Attitude_DL	3.33	.45	2.00	-.07 (.42)	-.08 (.82)	-.01	.05	.02	-.01	-.13*	.66**	1		
Attitude_S	3.45	.44	1.57	-.21 (.42)	-.96 (.82)	-.04	.03	-.02	-.05	-.10	.68**	.55**	1	
Attitude_IL	4.00	.33	1.19	.41 (.42)	-.57 (.82)	-.03	.04	-.13*	.02	-.07	.24**	.40**	.23**	1

Note: T = Technology; DL = Design Learning; S = Science; IL = Inquiry Learning

*p < .05, **p < .01

As shown in Table 5, results from the baseline model indicate that pupils' engagement, as rated by observation, increases over the four measurement occasions. An estimation of the random part of the baseline model shows that the variance in intercepts (σ_{cons}^2) at the class, pupil and time level is significant for the outcome (see Table 5). When we allowed the time variable to vary across classes, significant random slopes were detected, meaning that classes differ in their growth in pupils' engagement.

Therefore, it was possible to further explore whether teachers' attitudes towards S&T (teaching) are predictive for this growth in engagement over the four measurement rounds. With regard to the specific predictors in Models A[34] ('Engagement all measurement rounds' as dependent variable) we didn't find any of the attitudes to be predictive for the difference between classes in pupils' growth in engagement over all measurement rounds.

Figure 1. Models A

PREDICTOR VARIABLES		OUTCOME VARIABLE
Class level (level 3)		**Child level (level 2)**
*Teacher's attitudes		*Pupil's engagement
INTERACTING VARIABLES	----------------------------->	
Time level (level 1)		
*Time		

[34] In Figure 1 the tested Model is presented.

Table 5: Predicting Growth in Pupil's Engagement by Teachers' Attitudes towards S&T (teaching)

	Pupils' Engagement (all measurement occasions) (n = 1075)[a]																	
	Baseline Model (time not allowed to vary across classes)			Baseline Model (time allowed to vary across classes)			Model A1[b]			Model A2			Model A3			Model A4		
	B	SE		B	SE		B	SE	Δ R2	B	SE	Δ R2	B	SE	Δ R2	B	SE	Δ R2
Random parameters									.01			.00			.00			.00
Class level																		
σ_{cons}^2	.04**	.02		.07*	.03		.06*	.03		.07*	.03		.07*	.03		.07*	.03	
$\sigma_{cons,time}$				-.02*	.01		-.02	.01		-.02	.01		-.02*	.01		-.02*	.01	
σ_{time}^2				.01*	.01		.01*	.01		.01*	.01		.01*	.01		.01*	.01	
Pupil level																		
σ_{cons}^2	.07***	.02		.07***	.02		.07***	.02		.07***	.02		.07***	.02		.07***	.02	
Time level																		
σ_{cons}^2	.42***	.02		.40***	.02		.40***	.02		.40***	.02		.40***	.02		.40***	.02	
Deviance				2268.098			2265.38			2265.566								
			β			β			β			β			β			β
Fixed parameters																		
Intercept	3.12***	.05		3.12***	.06		3.13***	.06		3.13***	.06		3.12***	.06		3.12	.06	
Time[c]	.10***	.02	.17***	.10**	.03	.15***	.09**	.03	.14**	.09***	.03	.14***	.09***	.03	.14***	.09***	.03	.14***
Attitude T							.23	.14	.13									
Attitude T x Time							-.10	.07	-.10									
Attitude DL										.13	.14	.08						
Attitude DL x Time										-.09	.06	-.11						
Attitude S													.10	.14	.06			
Attitude S x Time													-.08	.06	-.09			
Attitude IL																.01	.19	.01
Attitude IL x Time																-.03	.09	-.03

[a] In the prediction of pupils' engagement (all measurement occasions), data were available on all predictor variables (attitudes) for 323 children, yielding (323 x 4 measurement occasions =) 1292 data points of information. Over the 4 measurement occasions, 217 pieces of information were missing on pupils' engagement (leading to n = 1075).

[b] In Models A (A1-A4), interaction effects of each attitude separately and the predictor time were added to the model.

[c] In this table the results are reported for measurement occasion 1.

*p < 0.05, **p < 0.01; *** p < 0.001

Relation between teachers' attitudes towards S&T (teaching)/ teacher-pupil and pupil-pupil interactions and pupil's growth in engagement

Next, to determine whether pupils grew in their engagement from pre to post, a paired samples t-test was conducted. This test showed that pupils' engagement in the post-measurement differs significantly from their engagement in the pre-measurement ($t\ (220) = 5.596$; $p = .000$). This made it possible to further investigate models with pupils' engagement growth (post-pre) as dependent variable. Next, in order to determine whether teachers' attitudes and the mean scores on CLASS domains/dimensions for each class over all measurement rounds function as predictors for pupils' engagement growth (post-pre), these variables were added in Models B, C and D. In Tables 6 and 7 descriptive statistics of the Models' variables and correlations of the predictor variables with the outcome variable can be found.

Table 6. Means, Standard Deviations, Range, Skewness and Kurtosis of Study Variables (Pupils' Engagement Growth (post-pre) as outcome)

	M	SD	Range	Skewness (SE)	Kurtosis (SE)
1.Engagement (post-pre)	.32	.85	5.25	.14(.17)	.48(.33)
2. Pos_Clim_mean_all_ occasions	4.90	.49	2.08	-.64(.43)	.34(.85)
3. Teach_Sens_mean_all_occasions	5.17	.44	1.88	-.46(.43)	-.16(.85)
4. Reg_Stud_Persp_mean_all_occasions	4.78	.25	1.25	-.81(.43)	2.58(.85)
5. Beh_Man_mean_all_occasions	5.67	.31	1.13	-1.30(.43)	1.17(.85)
6. Prod_mean_all_occasions	5.47	.26	1.25	-.51(.43)	.77(.85)
7. Neg_Clim_Rev_mean_all_occasions	6.58	.40	1.88	-2.10 (.43)	5.67(.85)
8. Cont_Und_mean_all_occasions	4.06	.60	2.38	-.40(.43)	.02(.85)
9. An_and_Inq_mean_all_occasions	4.50	.53	1.75	-.33(.43)	-1.03(.85)
10. Qual_of_Feedb_mean_all_occasions	4.31	.53	2.13	.36(.43)	-.20(.85)
11. Instr_Dial_mean_all_occasions	4.73	.41	1.63	.08(.43)	-.82(.85)
12. Em_Supp_mean_all_occasions	4.95	.29	1.32	-.99(.43)	1.19(.85)
13. Class_Org_mean_all_occasions	5.90	.23	1.07	-1.06(.43)	2.01(.85)
14. Instr_Supp_mean_all_occasions	4.58	.36	1.55	-.09(.43)	.55(.85)
15. Attitude_T	3.77	.40	1.93	-.68(.43)	1.24(.85)
16. Attitude_DL	3.34	.46	2.00	-.11(.43)	-.24(.85)
17. Attitude_S	3.44	.44	1.57	-.15(.43)	-1.03(.85)
18. Attitude_IL	4.00	.33	1.19	.41(.43)	-.63(.85)

Note: T = Technology; DL = Design Learning; S = Science; IL = Inquiry Learning

Table 7. Bivariate Correlations Between Study Variables (Pupils' Engagement Growth (post-pre) as outcome)

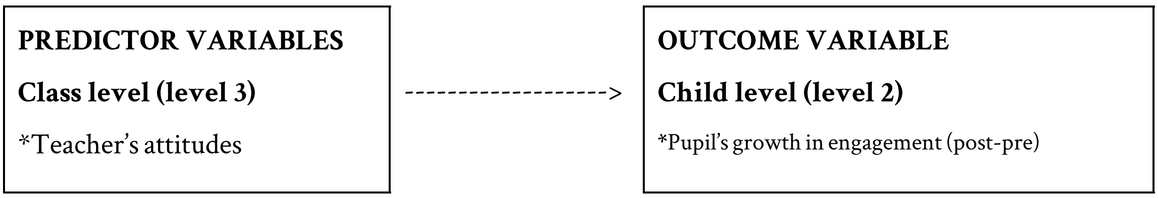
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Engagement (post-pre)	1																	
2. Pos_Clim_mean_all_ occasions	-.07	1																
3. Teach_Sens_mean_all_occasions	.14*	.45*	1															
4. Reg_Stud_Persp_mean_all_occasions	.05	.36	-.10	1														
5. Beh_Man_mean_all_occasions	.10	-.02	.14	-.19	1													
6. Prod_mean_all_occasions	-.16*	.20	.27	.13	.14	1												
7. Neg_Clim_Rev_mean_all_occasions	-.11	.63**	.37	-.01	.26	.19	1											
8. Cont_Und_mean_all_occasions	-.18**	.09	-.02	-.08	-.06	.38*	.31	1										
9. An_and_Inq_mean_all_occasions	-.03	-.06	.12	.18	.04	.18	.14	.59**	1									
10. Qual_of_Feedb_mean_all_occasions	-.03	.39*	.23	.23	.11	.25	.48**	.59**	.59**	1								
11. Instr_Dial_mean_all_occasions	.04	.27	.20	.48**	.18	.21	.45*	.21	.53**	.70**	1							
12. Em_Supp_mean_all_occasions	.05	.89**	.73**	.44*	.01	.29	.54**	.02	.07	.40*	.39*	1						
13. Class_Org_mean_all_occasions	-.08	.45*	.39*	-.04	.67**	.57**	.79**	.30	.17	.43*	.43*	.44*	1					
14. Instr_Supp_mean_all_occasions	-.05	.24	.23	.19	.14	.36	.47*	.76**	.83**	.89**	.72**	.30	.49**	1				
15. Attitude_T	-.15*	-.05	-.01	-.15	-.01	.04	.05	.07	.05	.07	.04	-.08	.05	.04	1			
16. Attitude_DL	-.14*	.06	-.04	-.05	.12	-.02	.14	-.05	-.06	.06	.05	.00	.13	-.01	.70**	1		
17. Attitude_S	-.16*	.23	.19	-.22	-.21	-.09	.26	.21	.21	.21	.12	.16	.03	.21	.68**	.55**	1	
18. Attitude_IL	.02	-.19	-.12	-.12	-.46*	-.26	-.37*	-.22	-.17	-.11	-.30	-.20	-.53**	-.28	.20	.37	.17	1

Note: T = Technology; DL = Design Learning; S = Science; IL = Inquiry Learning

*p < .05, **p < .01

The multilevel analyses show that some amount of variance is explained in pupils' growth in engagement (see Tables 8, 9 and 10)[35]. When the Emotional Support dimensions were simultaneously added to the baseline model as predictors (Model c1), the deviance of the total model reduced significantly ($X^2(3) = 7.91^*$). Model c1 shows that positive climate is negatively and teacher sensitivity positively related to pupils' growth in engagement when controlling for the other dimensions of the Emotional Support domain (see Table 9) or for positive climate separately (model d22, Table 10). A higher positive climate predicts a lower growth in pupils' engagement, and a higher sensitivity of the teacher resulted in a higher growth in engagement in the investigated primary school classes. Also when controlling for classroom organisation (model d9, Table 10), teacher sensitivity positively predicts pupils' growth in engagement. In the same line as the negative relationship found with positive climate, negative climate (reversed) also predicts pupil's growth in engagement negatively when controlling for teacher sensitivity (model d34). For both models (d22 and d34) the deviance reduced significantly ($X^2(2)= 6.24^*$ for the model with positive climate and $X^2(2)= 6.32^*$ for the model with Negative Climate). When controlling for each instructional support dimension separately and for the teacher sensitivity dimension separately, a higher content understanding of the teacher is associated with a lower growth in pupils' engagement (models d35, d61, d62, d63). Finally, when controlling for teacher sensitivity, productivity predicts pupil's growth in engagement in a negative way (models d33). The variance situated at class level is no longer significant for pupils' growth in engagement after adding the predictors in Models c1, d9, d22, d33, d34, d35, d61, d62 and d63. Although significant effects were found, the deviance of some models (models d9, d33, d35, d61, d62, d63) did not reduce significantly.

Figure 2. Models B



[35] Figures 2-4 provide a visual presentation of the tested models (summarised).

Figure 3. Models C

PREDICTOR VARIABLES		OUTCOME VARIABLE
Class level (level 3) *CLASS dimensions/CLASS domain **CONTROL VARIABLES** **Class level (level 3)** ***Other CLASS dimensions of the same domain/other CLASS domains**	------------------->	**Child level (level 2)** ***Pupil's growth in engagement (post-pre)**

Figure 4. Models D

PREDICTOR VARIABLES		OUTCOME VARIABLE
Class level (level 3) *CLASS dimensions/domains **CONTROL VARIABLES** **Class level (level 3)** ***Other CLASS dimensions/CLASS-dimensions not belonging to the domain**	------------------->	**Child level (level 2)** ***Pupil's growth in engagement (post-pre)**

Table 8: Predicting Pupil's Engagement Growth (post-pre) by Teachers' Attitudes Towards S&T (teaching)

	Pupil's Engagement Growth (post-pre) (n = 216)[a]														
	Baseline Model			Model B1			Model B2[b]			Model B3			Model B4		
	B	SE		B	SE	Δ R2	B	SE	Δ R2	B	SE	Δ R2	B	SE	Δ R2
Random parameters						.02			.02			.02			.00
Class level															
σ_{cons}^2	.11*	.05		.09	.05		.09	.05		.09	.05		.11*	.05	
Pupil level															
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	532.411			529.956			529.900			529.500			532.402		
			β			β			β			β			β
Fixed parameters															
Intercept	.32***	.09		1.55*	.77		1.30*	.61		1.40*	.62		.22	1.14	
Attitude T				-.33	.20	-.16									
Attitude DL							-.30	.18	-.16						
Attitude S										-.32	.18	-.17			
Attitude IL													.03	.29	.01

[a] In the prediction of pupils' engagement growth (post-pre), data were available on all predictor variables (attitudes and CLASS variables) for 303 children. 87 pieces of information were missing on pupils' engagement growth (leading to n=216).

[b]In Models B (B1-B4), each attitude was added to the model separately.

*p < 0.05, **p < 0.01; *** p < 0.001

Table 9: Predicting Pupil's Engagement Growth (post-pre) by Teacher-Pupil/Pupil-Pupil Interactions (Models C)

	Pupil's Engagement Growth (post-pre) (n = 216)[a]														
	Baseline Model		Model C1			Model C2[b]			Model C3			Model C4		
	B	SE	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2
Random parameters					.05			.04			.04			
Class level														
σ_{cons}^2	.11*	.05	.06	.04		.07	.04		.07	.04		.10*	.05	
Pupil level														
σ_{cons}^2	.62***	.06	.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	532.411		524.501			526.774			527.975			531.014		
					β			β			β			β
Fixed parameters														
Intercept	.32***	.09	.33***	.08		2.13	1.27		.35***	.08		.32***	.08	
Emotional Support[c]												.29	.34	.10
Positive Climate			-.48*	.20	-.28*									
Teacher Sensitivity			.54**	.20	.27**									
Regard for Student Perspectives			.48	.36	.14									
Classroom Organisation												-.37	.44	-.10
Behavior Management						.41	.25	.15						
Productivity						-.45	.29	-.14						
Negative Climate (rev.)						-.27	.19	-.13						
Instructional Support												-.08	.26	-.03
Content Understanding									-.39	.20	-.28			
Analysis and Inquiry									.10	.22	.06			
Quality of Feedback									.15	.26	.09			
Instructional Dialogue									-.01	.30	-.00			

[a] In the prediction of pupils' engagement growth (post-pre), data were available on all predictor variables (attitudes and CLASS variables) for 303 children. 87 pieces of information were missing on pupils' engagement growth (leading to n=216).

[b] In Models C (C1-C4), three models in which the CLASS dimensions of the same domain were added as well as a model in which the three CLASS domains were added were tested.

[c] CLASS domain and dimension scores with Pupil's Engagement Growth (post-pre) as dependent variable represent the mean of teachers' scores on that domain/dimension in the pre-, in-between the project-, and post-measurements. As some teachers had missing scores for one or more measurement occasions, for 23 teachers a mean score for the four measurement rounds could be calculated. A mean score was also calculated for teachers who had scores for three measurement occasions. These teachers were also included in the data, resulting in a total sample of 30 teachers.

*p < 0.05, **p < 0.01; *** p < 0.001

Table 10: Predicting Pupil's Engagement Growth (post-pre) by Teacher-Pupil/Pupil-Pupil Interactions (Models D)

Pupil's Engagement Growth (post-pre) .85
(n = 216)[a]

	Baseline Model		Model D1			Model D2[b]			Model D3			Model D4			Model D5			Model D6			Model D7		
	B	SE	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2
Random parameters					.01			.03			.02			.03			.01			.01			.00
Class level																							
σ_{cons}^2	.11*	.05	.10*	.05		.08	.05		.09	.05		.08	.04		.10*	.05		.10*	.05		.11*	.05	
Pupil level																							
σ_{cons}^2	.62***	.06	.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	532.411		531.404			529.949			529.193			528.982			532.128			531.936			532.264		
Fixed parameters					β			β						β						β			β
Intercept	.32***	.09	.33***	.08		.33***	.08		.32***	.08		.34***	.08		.32***	.09		.32***	.09		.32***	.09	
Emotional Support			.09	.31	.03	.22	.30	.08	.47	.36	.16	.14	.29	.05	.13	.32	.04	.19	.34	.06	.08	.33	.03
Behavior Management			.25	.26	.09																		
Productivity						-.49	.31	-.15															
Negative Climate (rev.)									-.42	.23	-.20												
Content Understanding												-.25	.13	-.18									
Analysis and Inquiry															-.07	.16	-.04						
Quality of Feedback																		-.10	.17	-.06			
Instructional Dialogue																					.04	.21	.20

	Model D8			Model D9			Model D10			Model D11			Model D12			Model D13			Model D14		
	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2
Random parameters			.01			.03			.01			.03			.01			.01			.01
Class level																					
σ_{cons}^2	.10*	.05		.08	.04		.10*	.05		.08	.05		.10*	.05		.10*	.05		.10*	.05	
Pupil level																					
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	531.259			527.987			531.455			529.136			531.720								
			β			β			β			β			β			β			β
Fixed parameters																					
Intercept				.33***	.08		.32***	.09		.34***	.08		.33***	.08		.32***	.08		.32***	.08	
Classroom Organisation	-.17	.40	-.05	-.59	.36	-.16	-.28	.36	-.08	-.10	.36	-.03	-.28	.37	-.08	-.28	.40	-.08	-.43	.40	-.12
Positive Climate	-.15	.21	-.09																		
Teacher Sensitivity				.40*	.20	.21*															
Regard for Student Perspectives							.22	.40	.06												
Content Understanding										-.24	.14	-.17									
Analysis and Inquiry													-.03	.17	-.02						
Quality of Feedback																-.01	.17	-.01			
Instructional Dialogue																			.17	.22	.08

	Model D15			Model D16[b]			Model D17			Model D18			Model D19			Model D20		
	B	SE	ΔR^2	B	SE		B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2
Random parameters			.01			.02			.01			.01			.02			.01
Class level																		
σ_{cons}^2	.10*	.05		.09	.05		.10*	.05		.10*	.05		.09	.05		.10*	.05	
Pupil level																		
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	531.318			529.520			531.495			530.855			530.462					
			β			β			β			β			β			β
Fixed parameters																		
Intercept	.33***	.08		.34***	.08		.32***	.09		.33***	.08		.33***	.08		.33***	.08	
Instructional Support	-.08	.23	-.03	-.22	.22	-.09	-.18	.23	-.08	-.18	.22	.09	-.01	.24	-.00	-.00	.25	-.00
Positive Climate	-.17	.19	-.10															
Teacher Sensitivity				.31	.19	.16												
Regard for Student Perspectives							.31	.41	.09									
Behavior Management										.29	.26	.11						
Productivity													-.43	.33	-.13			
Negative Climate (Rev.)																-.25	.23	-.12

	Model D22			Model D23			Model D24			Model D25			Model D26			Model D27			Model 28			Model 29			Model 30		
	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2
Random parameters			.04			.01			.01			.02			.01			.03			.01			.01			.01
Class level																											
σ_{cons}^2	.07	.04		.10*	.05		.10*	.05		.09	.05		.10*	.05		.08	.04		.10*	.05		.10*	.05		.10*	.05	
Pupil level																											
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	526.169			530.824			530.456			529.805						528.636			531.336			531.438			531.206		
			β			β			β			β			β			β			β			β			β
Fixed parameters																											
Intercept	.35***	.08		.32***	.09		.33***	.08		.34***	.08		.33***	.08		.35***	.08		.33***	.08		.33***	.08		.33***	.08	
Positive Climate	-.42*	.20	-.24*	-.21	.19	-.12	-.19	.18	-.11	-.15	.18	-.09	-.06	.24	-.03	-.14	.18	-.08	-.18	.19	-.10	-.18	.20	-.10	-.20	.19	-.12
Teacher Sensitivity	.49*	.20	.25*																								
Regard for Student Perspectives				.32	.41	.09																					
Behavior Management							.26	.26	.09																		
Productivity										-.41	.31	-.13															
Negative Climate (Rev.)													-.21	.26	-.10												
Content Understanding																-.23	.14	-.16									
Analysis and Inquiry																			-.05	.16	-.03						
Quality of Feedback																						-.01	.17	-.01			
Instructional Dialogue																									.10	.20	.05

	Model D31			Model D32			Model D33			Model D34			Model D35			Model D36			Model37			Model 38		
	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2
Random parameters			.02			.02			.04			.04			.04			.02			.02			.02
Class level																								
σ_{cons}^2	.09	.05		.09	.05		.06	.04		.07		.04	.07	.04		.09	.05		.09	.05		.09	.05	
Pupil level																								
σ_{cons}^2	.62***	.06		.62***	.06		.63***	.06		.62***		.06	.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	530.040			529.910			526.992			526.091			526.826			530.158			529.880					
			β			β			β			β						β			β			β
Fixed parameters																								
Intercept	-2.34	2.15		-2.03	1.59		1.70	1.65		.34***	.08		.35***	.08		.33***	.08		.33***	.08		.33***	.08	
Teacher Sensitivity	.27	.19	.14	.24	.19	.12	.34	.18	.18	.44*	.19	.23*	.28	.17	.14	.28	.19	.14	.30	.19	.16	.27	.19	.32
Regard for Student Perspectives	.26	.39	.07																					
Behavior Management				.20	.26	.07																		
Productivity							-.57*	.29	-.17*															
Negative Climate (Rev.)										-.44*	.20	-.21*												
Content Understanding													-.25*	.13	-.18*									
Analysis and Inquiry																-.09	.16	-.05						
Quality of Feedback																			-.12	.15	-.07			
Instructional Dialogue																						.01	.20	.00

	Model D39			Model D40			Model D41			Model D42			Model D43			Model D44			Model45		
	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2
Random parameters			.01			.03			.01			.03			.01			.01			.00
Class level																					
σ_{cons}^2	.10*	.05		.08	.05		.10*	.05		.08	.05		.10*	.05		.10*	.05		.11*	.05	
Pupil level																					
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	530.903			529.804			530.568			529.074			531.748			531.735			532.070		
			β			β			β			β			β						β
Fixed parameters																					
Intercept	.32***	.08		.32***	.08		.32***	.08		.34***	.08		.31***	.09		.32***	.09		.31***	.09	
Regard for Student Perspectives	.31	.40	.09	.32	.38	.09	.21	-.25	.06	.15	.38	.04	.31	.43	.09	.29	.42	.09	.24	.47	.07
Behavior Management	.29	.26	.11																		
Productivity				-.48	.30	-.15															
Negative Climate (Rev.)							-.25	.20	-.12												
Content Understanding										-.24	.14	-.17									
Analysis and Inquiry													-.10	.17	-.06						
Quality of Feedback																-.09	.16	-.06			
Instructional Dialogue																			.00	.23	.00

	Model D46			Model D47			Model D48			Model D49			Model D50			Model D51		
	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2
Random parameters			.03			.02			.03			.01			.01			.01
Class level																		
σ_{cons}^2	.08	.04		.09	.05		.08	.04		.10*	.05		.10*	.05		.10*	.05	
Pupil level																		
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	528.731			528.943			528.402			531.295			531.148			531.477		
			β			β			β			β			β			β
Fixed parameters																		
Intercept	.34***	.08		.33***	.08		.34***	.08		.33***	.08		.33***	.08		.33***	.08	
Behavior Management	.33	.25	.12	.36	.26	.13	.22	.24	.08	.26	.26	.09	.28	.26	.10	.25	.27	.09
Productivity	-.52	.30	-.08															
Negative Climate (Rev.)				-.33	.20	-.16												
Content Understanding							-.24	.13	-.17									
Analysis and Inquiry										-.07	.16	-.04						
Quality of Feedback													-.09	.15	-.06			
Instructional Dialogue																.02	.21	.01

	Model D52			Model D53			Model D54			Model D55			Model D56		
	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2	B	SE	ΔR^2
Random parameters			.02			.03		.02				.02			.03
Class level															
σ_{cons}^2	.09	.05		.08	.04		.09	.05		.09	.05		.08	.05	
Pupil level															
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	529.419			528.636			530.465			530.464			530.033		
			β			β			β			β			β
Fixed parameters															
Intercept	.33***	.08		.34***	.08		.33***	.08		.33***	.08		.33***	.08	
Productivity	-.38	.31	-.12	-.25	.33	-.08	-.44	.32	-.13	-.44	.32	-.13	-.49	.31	-.15
Negative Climate (Rev.)	-.20	.20	-.09												
Content Understanding				-.20	.15	-.14									
Analysis and Inquiry							.00	.16	.00						
Quality of Feedback										-.01	.15	-.01			
Instructional Dialogue													.13	.19	.06

	Model D57			Model D58			Model D59			Model D60		
	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2
Random parameters			.03			.01			.01			.02
Class level												
σ_{cons}^2	.08	.04		.10*	.05		.10*	.05		.09	.05	
Pupil level												
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06		.62***	.06	
Deviance	528.642			530.820			530.805			529.932		
			β			β			β			β
Fixed parameters												
Intercept	1.21*	.59		.43	.73		.19	.75		.32***	.08	
Negative Climate (Rev.)	-.15	.20	-.07	-.25	.20	-.12	-.27	.23	-.13	-.35	.22	-.16
Content Understanding	-.22	.14	-.16									
Analysis and Inquiry				-.02	.16	-.01						
Quality of Feedback							.03	.17	.02			
Instructional Dialogue										.21	.22	.10

	Model D61			Model D62			Model D63		
	B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2
Random parameters			.03			.03			.03
Class level									
σ_{cons}^2	.08	.04		.08	.04		.08	.04	
Pupil level									
σ_{cons}^2	.62***	.06		.62***	.06		.62***	.06	
Deviance	528.486			528.214			528.514		
			β			β			β
Fixed parameters									
Intercept	.34***	.08		.34***	.08		.34***	.08	
Content Understanding	-.33*	.16	-.23*	-.36*	.17	-.25*	-.28*	.14	-.20*
Analysis and Inquiry	.16	.18	.10						
Quality of Feedback				.19	.19	.12			
Instructional Dialogue							.16	.19	.08

	Model D64			Model D65		
	B	SE	Δ R^2	B	SE	Δ R^2
Random parameters			.00			.01
Class level						
σ_{cons}^2	.11*	.05		.10*	.05	
Pupil level						
σ_{cons}^2	.62***	.06		.62***	.06	
Deviance	532.219			531.852		
			β			β
Fixed parameters						
Intercept	.33***	.09		.32***	.08	
Analysis and Inquiry	-.02	.21	-.01	-.14	.21	-.09
Quality of Feedback	-.05	.20	-.03			
Instructional Dialogue				.17	.16	.08

	Model D66		
	B	SE	ΔR^2
Random parameters			.01
Class level	.10*	.05	
σ_{cons}^2			
Pupil level			
σ_{cons}^2	.62***	.06	
Deviance	531.576		
			β
Fixed parameters			
Intercept	.32***	.08	
Quality of Feedback	-.18	.21	-.11
Instructional Dialogue	.22	.28	.11

[a] In the prediction of pupils' engagement growth (post-pre), data were available on all predictor variables (attitudes and CLASS variables) for 303 children. 87 pieces of information were missing on pupils' engagement growth (leading to n=216).

[b] In models D (D1-D66), the domains and dimensions were added while controlling for each separate CLASS dimensions (in case of adding a domain there was not controlled for the dimensions of that particular domain).

[c] CLASS domain and dimension scores with Pupil's Engagement Growth (post-pre) as dependent variable represent the mean of teachers' scores on that domain/dimension in the pre-, in-between the project-, and post-measurements. As some teachers had missing scores for one or more measurement occasions, for 23 teachers a mean score for the four measurement rounds could be calculated. A mean score was also calculated for teachers who had scores for three measurement occasions. These teachers were also included in the data, resulting in a total sample of 30 teachers.

*p < 0.05, **p < 0.01; *** p < 0.001

Conclusions and discussion

In this study – via the first research question – we investigated the evolution in pupils' engagement over the course of the implementation of the project-based S&T learning environment Village@School, and the differences between schools/classes with regard to this possible growth. In particular, we were interested in the possible explanatory role of teacher's attitudes and the quality of interactions. This interest was represented in a second research question. During the implementation of the challenging Village@School project, teachers were supported by an introductory conference and ongoing workshops and coaching sessions.

The first conclusion of this study is that pupils' engagement grew over the course of the Village@School project, in which pupils were challenged to design and construct a miniature site in a period of minimum 10 weeks (*research question 1*). This is in line with previous studies like that of Mant et al. (2007), in which an increase in 10 to 11-year-old pupils' engagement was detected after a continuing professional development initiative for their teachers which focused on creating stimulating science lessons. The Village@School project itself may have caused this growth in engagement. For most pupils, Village@School is an innovative project because of the high level of autonomy that is given to them. Pupils are often not used to this room for initiative, and must learn to make their way. The organisational format – consisting of a challenging goal which requires elaborate planning and division of tasks (doing research, making constructions, doing the accounting, searching for materials etc.) – contains a lot of opportunities for pupils to be engaged, and pupils appear to grow in making the most of these opportunities throughout the project. As pupils gradually became accustomed to the higher level of autonomy, their self-regulation may have increased, leading to empowering, successful actions which were visible in their engagement. Interest and enjoyment of the activity for its own sake are the basis for internally valued, regulated and motivated behaviour (Stefanou, Perencevich, DiCintio, & Turner, 2004). Other studies have also detected a positive change; not in pupils' engagement, but in their related S&T attitudes, as a result of participation in science programmes that are inquiry-oriented, involve hands-on exploration with meaningful materials, promote classroom discussion, and make science exciting (Smith, 2015; Vargas-Gomez & Yager, 1987). Important to remark is that these attitudes are even more robust because they refer to "a general and enduring positive or negative feeling about science" (Koballa & Crawley, 1985, p. 222), whereas engagement is more connected to the specific educational activities provided (Fredricks et al., 2004). Nevertheless, we cannot conclude that the whole intervention was effective, as an important limitation of this study is that it did not involve a control group.

The second conclusion of this study, also with regard to *research question 1*, is that the group of classes that participated in the Village@School intervention showed substantial differences with regard to their growth in engagement. Therefore, we investigated whether these differences were related to teachers' attitudes towards S&T (teaching) and the quality of the interactions (*research question 2*).

With regard to the interactions, we first found that when controlling for the other Emotional Support dimensions – Positive Climate and Regard for Student Perspectives – a higher teacher sensitivity results in a higher growth in pupils' engagement, as measured after the intervention. A sensitive teacher is aware of students' problems and needs, and is responsive by providing reassurance, giving stimulating impulses for re-engagement, and by solving their issues and questions (Hamre & Pianta, 2007). When the teacher is sensitive, students also feel comfortable in participating, taking risks and asking questions. While teacher sensitivity is often underestimated in S&T learning environments, as the emphasis rests on thought- and talk-provoking interactions (Damhuis & De Blauw, 2011), we have found that there are differences in sensitivity and that they matter where growth in pupils' engagement is concerned. This finding is not surprising, as a lot of pupils were not used to a challenging project like Village@School in which they 'suddenly' had to take initiative and work together to reach the goal, met difficulties, and had to be encouraged to think about alternative solutions for a problem.

Secondly, two surprising relationships were found in the prediction of pupils' engagement: a more positive climate and more stimulation for content understanding via content, materials etc. resulted in a lower rather than a higher growth in pupils' engagement, when controlling for the other Emotional Support dimensions (in the case of positive climate) and for Teacher Sensitivity and three other Instructional Support dimensions (in the case of content understanding). In general, students' perceptions of classroom atmosphere and teacher enthusiasm have been mentioned as factors in the literature on motivation (Clark, 1999; Meyer & Turner, 2002; Patrick, Hisley, & Kempler, 2000). Our first hypothesis is that in classes with a positive climate, teachers were already conducting different activities or projects in which pupils had to work together. As such, pupils may have been more used to taking initiative and cooperating, which could explain their lower growth in engagement throughout open-ended S&T activities. Another hypothesis for the relation with Positive Climate could be that in more open-ended learning environments, pupils who have more positive relationships with peers may be more easily distracted from the materials and content because they are eager to engage in social conversations with friends. It is possible that room for initiative does not work for every pupil, which may make him/her choose other activities such as chatting with friends. Finally, it is important to notice that in our study none of the classes scored 'low' on the Positive Climate dimension. Only two of them had a mean score lower than 4; the last being a mid-score in the

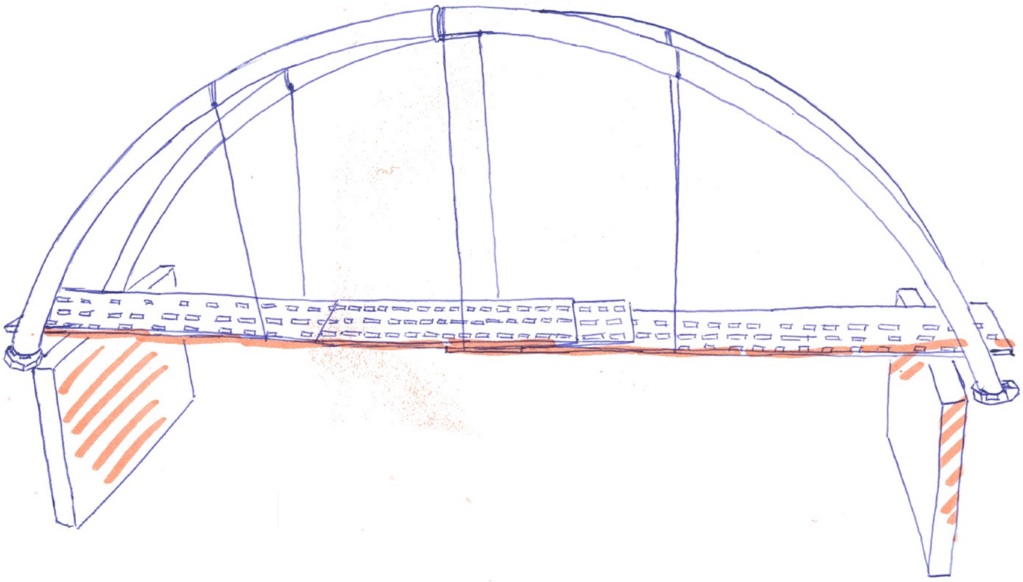
CLASS scoring system (Hamre & Pianta, 2007). As such, we cannot conclude that a *low* positive climate contributed to a higher growth in engagement .

With regard to content understanding, at first sight one would expect a higher or increasing level of engagement in pupils when teachers provide more realistic materials, content and questions that stimulate pupils' understanding. Nevertheless, we reiterate that pupils' engagement, as well as having a cognitive and behavioural component, also consists of an emotional component. As such, providing more complex content and materials may not automatically result in being more interested in and enjoying S&T activities. Moreover, further analyses in the pre-measurement showed that when controlling for productivity in the classroom, content understanding has a positive effect on pupils' engagement. This may indicate a ceiling effect, meaning that it may be difficult for engagement to grow because of a higher content understanding.

Surprisingly, no relationships was found with dimensions that, from a theoretical point of view, are seen as relevant for these learning environments. In particular, this applies to the dimensions Regard for Student Perspectives and Analysis and Inquiry . These dimensions evaluate whether the teacher gives autonomy and responsibility to pupils, connects content to pupils' reality and provides opportunities for collaboration and independent inquiry. Potvin and Hasni (2014) showed that such aspects are fruitful for the affective outcomes in students. However, because the same learning environment promoting these dimensions was implemented in all classes and all teachers were coached throughout the trajectory in finding a good balance between giving autonomy and intervening, the differences between classes concerning these dimensions were not that large; meaning that it may have made it difficult to find such effects in our sample. Moreover, open-ended S&T learning environments, and also the project-based learning environment Village@School, may have an internal structure that functions without or even in spite of (poor) teacher interventions.

As discussed, an important limitation of the experimental design formulated to answer *research question 1* is the lack of a control group. Despite this limitation, this study is unique in three respects. Firstly, fine-grained observation instruments were used to develop a profound understanding of pupils' engagement and the role of interactions in open-ended S&T learning environments. With the LIS-P (Laevers, 2011) and the CLASS Upper Primary (Pianta et al., 2012), we were able to measure the quality of interactions and the level of pupils' engagement on the basis of their behaviour in the S&T learning environment itself. In most studies, data related to pupils' engagement is gathered by asking pupils about their perceptions of their own engagement. By making observations, we could get a clear picture of what the learning environment realises in teachers and pupils. Secondly, intensive observations were conducted in an exceptionally large sample of classes. Finally, interactions in open-ended S&T learning environments had not previously been studied from a broader perspective which includes emotional, organisational and instructional aspects of interactions. As such, we tried to tackle the problem discussed by several researchers (e.g. Krajcik et al., 1998) regarding the lack of insight into how teachers realise high-quality open-ended S&T learning environments.

To conclude, this study has implications for the design and implementation of open-ended S&T learning environments in primary schools. Firstly, for pupils' growth in engagement it is important that pupils know they have a teacher they can rely on when confronted with problems and difficulties in these challenging S&T learning environments. One can suppose that a sensitive teacher is even more important in these student-directed learning environments in comparison to more teacher-directed learning environments. Secondly, a higher positive climate and level of content understanding do not appear to automatically contribute to pupils' growth in engagement. A lot of questions remain unanswered regarding this topic. Does a higher positive climate go together with more cooperative classroom activities, meaning that pupils grow less with regard to their engagement as a consequence of the high positive climate? Does the stimulation of a deep level of understanding via materials and content not automatically result in a more positive engagement, motivation, attitudes and interest because of the affective nature of these outcomes? Finally, perhaps the form – the way in which S&T is provided – is the decisive factor for such outcomes. The form of the activity rather than content topic and learning goals decide the interestingness of an instructional episode (Swarat, Ortony, & Revelle, 2012). With its inquiry- and problem-based nature, a project like Village@School is ambitious, as it presupposes that pupils are able to accept responsibility and take charge of solving problems and their learning. Therefore this project belongs among those interventions which have been shown to have positive effects on students' interest, motivation and attitudes (Potvin & Hasni, 2014). In future research, an experimental design in which a control group is used should verify this.



Study 3: An exploration of teachers' competence profile in science and technology (S&T): Its relation with the implementation of a challenging learning environment and opportunities for growth

Submitted

Keywords: elementary/primary school, teachers' competence profile towards science and technology (teaching), classroom interactions, inquiry- and design- based learning environments

Abstract

The relation between primary school teachers' attitudes towards science and technology (S&T) and the quality of the inquiry- and design-based learning environments provided by teachers is often investigated. In this study, the first point of interest is the relation between elementary school teachers' initial competence profile towards S&T(teaching) – their attitudes and their teacher style in a standardised S&T-assignment – and their teacher style while implementing the challenging Village@School project. Next, the study includes a systematic exploration of how teachers' attitudes and teacher style in S&T activities may be changed in a positive way. Results showed a positive association between the initial emotional support provided and the quality of emotional support during Village@School. Classroom organisation and the attitude towards inquiry learning were negatively related to emotional support. We concluded that the way in which these effects work should be further investigated. Finally, after the project, teachers didn't grow in their attitudes throughout the project, but practice showed better interactions involving Regard for Student Perspectives and Quality of Feedback. Surprisingly, there was a growth in Negative Climate (also during Village@School). The results provide evidence for the hypothesis that a training focusing on the competence of teachers instead of on a certain proven method or content, helps to develop a qualitative way of interacting in open-ended S&T learning environments.

Introduction

Inquiry- and design-based learning environments have proven to be important for students' engagement in the fields of science and technology (S&T) as well as for the development of their competences in these fields. While in inquiry-based learning environments, students primarily engage in scientifically oriented questions, find evidence and come up with explanations (National Research Council, 1996), in design-based learning environments (Fortus et al., 2004) the process of making a working product, of 'designing' a technological solution (NDET, 2006, in Hansen, 2010; Roth, 2001), is more prominent. A combination of inquiry and design is often established, usually in project-based learning environments (Barak & Doppelt, 2000). Typically, these inquiry-, design- or project-based learning environments are open-ended because they are characterised by a teacher who gives a large amount of initiative to students but who still maintains an active role (Hakkarainen, 2009; Kolodner, 2001). Because this role is different from traditional teaching, acting in these open-ended S&T learning environments is not self-evident for teachers (Veermans, Lallimo, & Hakkarainen, 2005). Therefore, it is important to investigate those characteristics of teachers that may promote high-quality interactions in these learning environments, which are important for pupils' engagement in making sense of phenomena or events in the natural world (Damhuis & De Blauw, 2011). Different studies have already shown that teachers' knowledge, skills and attitudes are related to their actual teaching of inquiry and design (Jones & Carter, 2007; Munck, 2007; Rohaan, Taconis & Jochems, 2012). Especially in primary education, teachers are often not specifically prepared to teach S&T, which may result in negative teacher attitudes (Cobern & Loving, 2002). A first aim of this intervention study is to investigate the relation between both Flemish and Dutch elementary school teachers' initial competence profile – as comprised of their attitudes towards S&T (teaching) and their teacher style in open-ended S&T activities – and their teacher style during the actual implementation of a challenging project-based S&T-learning environment, 'Village@School'. Teacher style is conceptualized as a general stance of the teacher towards his/her students (Sierens et al., 2006; Sweertvaegher, 2008), which is often understood in the interactions that occur between a teacher and his/her students (de Kruif, McWilliam, Ridley, & Wakely, 2000; Laevers & Heylen, 2013). Secondly, we will systematically explore whether the implementation of Village@School has an impact on teachers' initial competence profile, in terms of a positive change in attitudes towards S&T (teaching) and teacher style.

Theoretical framework

Interactions in open-ended S&T learning environments

Creating open-ended S&T learning environments starts with a rich offer of materials. There is however consensus in the literature that the material setting alone is not sufficient in order to engage students. Teacher style is important as it unfolds in classroom interactions, both between the teacher and the students and among the students. In the - compared to technology - more studied field of science researchers have focused on the concept of classroom discourse when investigating the processes of interactions in the classroom (Chin, 2007; Erdogan & Campbell, 2008; Hackling et al., 2011; Nystrand, et al., 2003; Reinsvold & Cochran, 2012; Scott et al., 2006; Smart & Marshall, 2013). In discourse analysis, different methods are used to analyse the written and spoken language of these interactions (Mercer, 2010; Nystrand et al., 2003). While these studies have their value for the investigation of interactions, it is important to take a broader perspective by studying also the more emotional, organisational and instructional aspects of classroom interactions that are part of or closely related to this classroom discourse (e.g. a positive climate which is characterized by warm relationships between teachers and students and among students; a deep understanding of content, via provided content and materials but also revealed in dialogues between students; behaviour management etc. (Pianta et al., 2012)) too. Large-scale studies, also in other fields than S&T, provided evidence for the existence of the three above-mentioned domains of interactions (Hamre, Pianta, Mashburn, & Downer, 2007): emotional support, classroom organisation, and instructional support (Hamre & Pianta, 2007; Hamre et al., 2007; Pianta et al., 2008).

In the literature on S&T education, the aspects that belong to the instructional support domain seem to have a higher value than the aspects belonging to the emotional support and classroom organisation domains. With regard to instructional support interactions, peer scaffolding in which students are able to instruct, pose questions, and give feedback to each other is highly valued, as in these open-ended S&T learning environments collaboration has a central role (Gnadinger, 2008). Students are better motivated and engaged when their views are sought and valued through dialogic teaching (Alexander, 2008; Mercer, Dawes & Staarman, 2009), which is characterized by an "active, influential and sustained participation of pupils in classroom talk" (Mercer et al., 2009, p. 354). While teachers no longer take on their traditional role, their role is still an active one in these learning environments. First, they should create the opportunities for collaboration and student centred discourse, by promoting those practices in which students utilise their own ideas and strengthen their own community (Viilo et al., 2011). Secondly, they should intervene actively in order to stimulate students' mental activity. The use of authentic questions – i.e. questions that do not have pre-specified answers –, higher-level questions and, especially, student-generated questions, are particularly important because they engage students (Nystrand et al., 2003). To develop students' competences via classroom discourse, a teacher needs to be able to integrate students' everyday experiences and make it possible to share these experiences with the whole learning community (Viilo et al., 2011). Above the frequently discussed guiding and scaffolding of students' processes by teachers, some authors point to the responsiveness of the teacher (Roth, 1998a), which is an aspect of the provided emotional support. In our opinion, this aspect as well as the classroom climate has been given little attention in the literature on S&T education. Finally, only some authors point to the importance of classroom management in open-ended S&T learning environments (Jacobowitz, 1997; Lawson, 1995).

Primary school teachers' attitudes towards S&T and its teaching

Unfolding classroom interactions in inquiry- and design-based learning environments may be a challenging task for primary school teachers. Studies show that teachers with less positive attitudes towards S&T (teaching) rely more on standardised methods and top-down instruction (Appleton & Kindt, 1999; Harlen & Holroyd, 1997; Jarvis & Pell, 2004; Plonczak, 2008). Especially in primary education - in which teachers are often not specifically prepared to teach science - this is a problem (Cobern & Loving, 2002). Therefore, one can suppose that teachers with positive attitudes towards S&T (teaching) are better able to cope with this challenge in comparison to others having less positive or even negative attitudes.

In general, attitudes are a summary evaluation of a psychological object in terms of favourable or unfavourable attribute dimensions such as good/bad or positive/negative (Ajzen, 2001; Eagly & Chaiken, 1993). An attitude is "a learned disposition to respond in a consistently favourable or unfavourable manner with respect to a given object" (Fishbein & Ajzen, 1975, in Young, 1998, p. 97). The most recent review concerning primary school teachers' attitudes towards the object of science[36] of Van Aalderen-Smeets, Walma van der Molen and Asma (2011) is the starting point for our theoretical background concerning primary school teachers' attitudes towards S&T. On the basis of different studies (Barmby et al., 2008; Bennett et al., 2001; Coulson, 1992; Osborne et al., 2003; Pajares, 1992) Van Aalderen-Smeets et al. (2011) conclude that the concept of attitude is often poorly articulated. Many studies provide incomplete definitions (or no definition at all) for the construct of attitude, fail to explicate the dimensions of attitude that they distinguish, and/or do not distinguish between attitudes and other related concepts (e.g. beliefs,

[36] Science in this study refers to both science and technology.

opinions, motivation) (Van Aalderen-Smeets et al., 2011). In spite of this, Van Aalderen-Smeets et al. (2011) found in the literature two recurring distinctions in the concept of teachers' attitudes towards S&T.

First, a distinction can be made between teachers' personal attitudes towards the disciplines of S&T as a citizen, independent of their profession, and their professional attitudes with regard to teaching science[37] and technology (Oberon, 2009; Van Aalderen-Smeets et al., 2011; Wilkins, 2008). For instance, a teacher's general interest in or affect towards science is part of his/her personal attitude, whereas the teacher's professional attitude, involves for example feelings of joy or anxiety with regard to teaching S&T.

Secondly, in the general literature on attitudes the construct is often divided into three dimensions: affective, cognitive, and behavioural (Eagly & Chaiken, 1993; Katz & Stotland, 1959; Klop & Severiens, 2007; Rosenberg & Hovland, 1960). Van Aalderen-Smeets et al. (2011) showed the existence of different attributes of each of these dimensions, and we selected relevant attributes in light of this study, which we will explain hereafter. These components are quite similar for the personal and the professional attitude of teachers towards S&T. The attributes were the same as the ones selected by Walma van der Molen (2007), who developed an instrument to assess attitude for Dutch teachers.

The affective dimension of a teacher's personal and professional attitude towards S&T consists of feelings and moods that he/she experiences in relation to these domains (Van Aalderen-Smeets et al., 2011). As often conceived in the literature, whether a teacher feels attracted by S&T (Palmer, 2004; Young, 1998) and the general enjoyment in teaching (Johnston & Ahtee, 2006; Ramey-Gassert, Schroyer, & Staver, 1996) are part of this dimension.

The cognitive dimension consists of the evaluative thoughts and beliefs towards S&T and its teaching. First, this dimension contains the perceived relevance or importance of science, which refers to "the extent to which people consider science relevant or important for their personal lives, for society, for prosperity, or for health" (Van Aalderen-Smeets et al., 2011, p.164). Teachers' perceptions of the relevance or importance of teaching science in school also belong to this dimension (Appleton & Kindt, 1999; Carleton, Fitch, & Krockover, 2008; Cobern & Loving, 2002; Johnston & Ahtee, 2006; Liang & Gabel, 2005). Secondly, the cognitive dimension relates to the perceived difficulty of S&T, which refers to the thoughts and beliefs of teachers concerning the general difficulty of S&T relative to other fields of study, and their

[37] As discussed in the introduction of this study, learning environments in which science is prominent are often called inquiry-based learning environments; learning environments in which technology is prominent are often called design-based learning environments.

perceptions about the difficulty of its teaching (Harlen & Holroyd, 1997; Johnston & Ahtee, 2006; Liang & Gabel, 2005).[38]

The behavioural dimension constitutes the behavioural responses or actions of a person when confronted with S&T and its teaching. This response can be either overt (with the person actually acting out the behavioural response or action) or covert (with the person intending to act out the behaviour, although the action has yet to take place). Though this dimension is seen as part of the attitude concept, Van Aalderen-Smeets et al. (2011) did not find any articles concerning primary school teachers' attitudes that reported about having measured behaviour related to S&T in respondents' daily lives or measured behavioural intention to engage in activities related to S&T. However, according to the authors a few studies investigated the behavioural component of attitude towards teaching science (Appleton & Kindt, 1999; Goodrum, Hackling & Rennie, 2001; Haney et al., 1996; Palmer, 2001; Yates & Goodrum, 1990).

In their recent review, Van Aalderen-Smeets et al. (2011) proposed a new theoretical framework for primary school teachers' attitudes towards science, in which a somewhat different perspective on the three parallel components is provided. In this framework they appeal to the Theory of Planned Behavior (Ajzen & Fishbein, 1980) in which behavioural intention is viewed as a direct outcome of the cognitive and affective dimension of attitudes, and not as a component of attitude itself. Furthermore, in the affective dimension they also distinguish 'anxiety' and a new dimension, 'perceived control', which consists of self-efficacy and context dependency, was added.

Relation between teachers' competence profile and teacher style in the project-based learning environment Village@School

In this study we investigate the relation between teachers' competence profiles with regard to S&T and their teacher style in the project-based learning environment Village@School. A teacher's competence profile can be described as the combination of the essential elements of professional competence, namely the knowledge, attitudes, and skills required for effective performance in a job (du Chatenier, Verstegen, Biemans, Mulder, & Otma, 2010; Mulder, 2001). In this study, we are interested in two aspects of a teacher's profile: his/her attitudes towards S&T (teaching) and his/her teacher style. As mentioned above, an indicator for teacher style is the quality of the interactions. This is the indicator the current study will focus on. We aim to get insight into the pattern of interactions in a classroom during open-ended S&T activities.

[38] Besides these different aspects, Van Aalderen-Smeets et al. (2011) also distinguish the thoughts and beliefs with regard to gender roles in S&T (e.g. the perception that men are better in understanding S&T than women).

In order to do so, the disposition of the teacher to have high-quality interactions with pupils during open-ended S&T activities is evaluated. Because the teacher via his/her interactions determines interactions among pupils too (Pianta et al., 2012), pupil-pupil interactions cannot be seen disconnected from teacher-pupil interactions.

The connection between teachers' attitudes towards S&T (teaching) and the frequency or quality with which learning activities in the domains of S&T are conducted has already been investigated extensively (e.g. Rohaan et al., 2010). Moreover, the literature shows that teachers' attitudes, knowledge and skills have an influence on the way in which new curricula are implemented (Krajcik, Blumenfeld, Marx, & Soloway, 1994; Rogers et al., 2011). However, most studies have focused on the relation between primary school teachers' attitudes towards S&T (teaching) and the quality of the realised learning environment, but have disregarded the relation between the teacher style in open-ended S&T classes and the teacher style in a new open-ended S&T learning environment.

Choi and Ramsey (2010) investigated 14 primary school teachers enrolled in a three-hour, elementary science methods course that emphasised teaching science as inquiry. Attitudes were measured via the Revised Science Attitude Scale (Bitner, 1994; Thompson & Shrigley, 1986) with items that evaluated enjoyment, importance, fear/comfort of teaching inquiry, difficulty, and interest. To evaluate whether teachers were able to implement inquiry-based activities after the course, semi-structured interviews, lesson plans, and written reflections of the participants were gathered and one lesson created by the teachers was observed. Finally, the Reformed Teaching Observation Protocol (RTOP), an observation instrument to observe the quality of science and mathematics education, was used (Lakshmanan, Heath, Perlmutter, & Elders, 2011; Piburn & Sawada, 2000; Sawada et al., 2000). If teachers felt comfortable teaching science and using inquiry-instruction, they were more likely to carry out inquiry-instruction in their lessons (Choi & Ramsey, 2010).

In the study by Choi and Ramsey (2010), classroom interactions were measured via the subscale 'communicative interactions' of the RTOP, as the authors' intention was to measure the quality of the inquiry-based learning environment in general. These 'communicative interactions' are similar to the concept of classroom discourse that we discussed earlier, and thus limited to the communication of students' ideas with each other or with the teacher.

In our operationalisation of teacher style, it would be possible to just focus on one aspect or to analyse the detailed conversations between teachers and pupils or among pupils, as is done in classroom discourse analysis (Mercer, 2010; Nystrand et al., 2003). Instead, a choice is made to evaluate different characteristics of the interactions. For example, the extent to which teachers take students' ideas into consideration is operationalised when studying classroom interactions. To do that, we make use of the Classroom Assessment Scoring System (CLASS) Upper Elementary (Hamre & Pianta, 2007), a domain-general instrument that aims to capture classroom interactions in all its facets. In this instrument not only the quality of teacher-student interactions but also the quality of interactions among students is evaluated (Pianta et al., 2012). To be able to measure the style of the teacher in open-ended S&T activities in general two standardised S&T-assignments were developed (in the Methods section light is shed on these assignments).

For the purposes of this study, we were interested in teacher style over the course of the implementation of the Village@School project. In this project, pupils were challenged to build a miniature site on a standard plate (1.22 by 2.44 meters) in the course of at least 20 two-hour sessions spread over at least 10 weeks. To integrate in the design and construction as many applications as possible, all conceivable 'problems' that were relevant from a technological viewpoint needed to be 'discovered' by the students themselves, solutions had to be found, choices had to be made and constructions needed to be built in a context of cooperative learning. The aim of the project was for teachers to give a large amount of autonomy to pupils, while still having an active role themselves. While teachers in a certain way 'received' the design of this learning environment, the way in which they dealt with the often 'new' characteristics of this learning context via their classroom interactions is interesting to investigate. Therefore, it is plausible to expect that more positive attitudes towards S&T (teaching) and/or a better teacher style in open-ended S&T activities before the project, are positively related to the teacher style in the innovative project-based learning environment Village@School.

Altogether, this results in the first and main research question of this study: Is there a relation between the teacher's initial competence profile and the way the Village@School project is implemented in the classroom by the teacher? We hypothesised that the better the teacher's competence profile before the implementation of the Village@School project, the better the teacher style during the Village@School project, as measured by the CLASS. The two aspects of the teacher's competence profile, both separately and together, will be investigated in relation to the teacher style during Village@School.

Changing teachers' competence profile

Today, a growing number of researchers widely recognise the need to better prepare teachers to accomplish meaningful science learning for students (Gess-Newsome, Southerland, Johnson, & Woodbury, 2003; NRC, 1996; Stuart & Thurlow, 2000; Watters & Ginns, 2000; Weld & Funk, 2005). Various studies (e.g., Harlen & Holroyd, 1997; Palmer, 2004; Shrigley, 1983; Trumper, 1998) have shown a generally low level of scientific and technological literacy among pre-service and in-service primary school teachers, and these teachers generally tend to have negative attitudes toward science.

At first, the core of the problem, that is primary school teachers' negative attitudes, needs to be tackled. This is important because research shows that these attitudes are predictive of their intention to teach science in the classroom (Haney et al., 1996; van Aalderen-Smeets & Walma van der Molen, 2013) as well as their classroom practices when they teach science (Haney, Lumpe, Czerniak, & Egan, 2002, in Van Aalderen-Smeets et al., 2013). Moreover, they are less able to stimulate a positive attitude towards science in their students (Jarvis & Pell, 2004; Van Driel al., 2001). In the literature, a lot of intervention studies – mainly in the domain of science – have been published that recognise the relevance of the process of science and inquiry-based approaches, i.e. cooperative learning and open-ended investigations in changing teachers' attitudes (e.g., Liang & Gabel, 2005; Martin-Dunlop & Fraser, 2008).

Secondly and next to the change in attitudes, there is also a need to improve teacher practices with regard to S&T, in order for teachers to teach science effectively (Murphy & Smith, 2012; Smith, 2014). Teacher professional development can influence teacher learning and classroom practice (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009). In the literature, there is a consensus that effective professional development programs should address particular needs and fit into the context in which they will be implemented (Smith, 2015). Furthermore, on-going reflections and support help to enhance teacher learning and sustain change toward adopting new teaching practices (Bell & Gilbert, 1994; Borko, 2004).

The implementation of the project-based learning environment Village@School and the training provided to conduct this implementation, aim to foster positive attitudes in participating teachers as well as improving teacher style in S&T activities. Before the start of the project the teachers attended a conference and a workshop, and in the course of the project they were offered a second workshop and two coaching sessions. While the coaching sessions were individual, and based on teachers' individual needs, the conference and workshops provided opportunities for teachers to interact with each other. Several studies highlight the importance of thoughtful reflection on instructional practice, of non-formal learning through interactions with colleagues and of formal learning opportunities (Brouwer & Korthagen, 2005; Park & Oliver, 2008; Van Driel, 2010; Zembal-Saul, Krajcik, & Blumenfeld, 2002). Next to investigating the growth in teacher style before and after Village@School, we also wanted to gain insight into what happens during the project itself with regard to this teacher style. Therefore, the growth in teacher style during Village@School was investigated as well.

The resulting second research question is as follows: Is there a positive evolution in the teacher's competence profile throughout and after the implementation of the Village@School project? First, the change in teachers' attitudes towards S&T (teaching) and teacher style before and after the implementation of the project was evaluated. Second, we investigated the evolution in the teacher's style over the course of the project. Because of their involvement in the project and the support provided, the hypothesis is that there will be an improvement in teachers' competence profile.

Method

Participants

Initially, 34 primary school teachers within eighteen schools volunteered to participate in this study. Four of these schools were located in the Netherlands (8 Dutch teachers), and fourteen in Belgium (26 Belgian teachers). The participating teachers taught in the 3^{rd} grade (3 groups of pupils), 4^{th} grade (7 groups of pupils), 5^{th} grade (16 groups of pupils), 6^{th} grade (16 groups of pupils), or in a combination of two consecutive grades. Four classes contained both 5^{th} and 6^{th} grade pupils; three classes had both 4^{th} and 5^{th} grade pupils and three others consisted of both a 3^{rd} and 4^{th} grade pupils. Two of these mixed classes - a 3^{rd} and 4^{th} grade and a 5^{th} and 6^{th} grade - belong to a school for highly gifted students. Shortly after the pre-measurement, in which we collected the data with regard to teachers' competence profile, and before the actual start of the Village@School project, five teachers dropped out of the study. We were able to replace them with two additional teachers, but for them not all pre-measurement data was available. This resulted in a new sample of 31 teachers. Two teachers from the same school took part in the study, except for two schools from which only one teacher participated[39] and three schools from which three teachers were engaged, the last because of the replacement of teachers[40]). Of all teachers, 26 teachers were female and 10 teachers were male. For the analyses conducted to answer the first research question, data was available for both attitudes towards S&T (teaching) and teacher style for 29 teachers. To determine the evolution in teachers' attitudes towards S&T (teaching) and teacher style, 34[41] teachers were involved in the analyses.

Because of practicability reasons, data was collected in two waves. The first wave (involving 19 teachers) started in November 2013 and ran until June 2014; the second wave (involving 17 teachers) in September 2014 and lasted until March 2015. In a pre- and post-measurement both teachers' initial attitudes and teacher style were measured. During the intervention teacher style was measured again.

[39] From one of these schools the participating teacher dropped out.

[40] In one school, one of two teachers dropped out and was replaced; in another school the two teachers dropped out and one replacing teacher (teaching the two classes of pupils) was involved; in still another school the only participating teacher dropped out in the first wave (see further) and in the second wave two other teachers of the same school participated in the study.

[41] For one of these teachers no data were available with regard to both attitudes and teacher style; for another one only data concerning the attitudes in the pre-measurement was available.

Data collection

Teachers' competence profile

Attitude questionnaire

A questionnaire developed by Oberon (2011) was administered to determine teachers' attitudes with regard to S&T and their teaching in the field. In line with the literature, the cognitive, affective and behavioural dimensions of attitudes towards S&T (teaching) are represented in the questionnaire. The questionnaire consists of a large part of items from the VTB-monitor (Walma van der Molen, 2007). The questionnaire contains 8 subscales for teachers' personal attitudes (4 for science and 4 for technology) and 8 subscales for their professional attitudes towards teaching S&T (4 for inquiry[42] and 4 for design), of which the items have to be rated on a 5-point Likert scale ('totally don't agree', 'don't agree', 'neutral', 'agree', 'totally agree'). The theoretically distinguished dimensions in the attitude concept fall apart in these scales. First, the affective dimension is evaluated via the 'pleasure' scale (e.g. 'I find technology interesting'; 'I like it more to leave children to their own devices, rather than that I tell them precisely how something works')[43]. Second, the cognitive dimension is operationalised on two scales: the 'difficulty' scale (e.g. 'I find doing inquiry hard'; 'For a teacher it is hard to find an appropriate problem that may be the starting point for design learning')[44] and the 'importance' scale (e.g. 'Technology is important for society'; 'Already in primary school children have to think in an inquiry way')[45]. The behavioural dimension finally is also represented via one scale, the 'intention future' scale (e.g. 'I like reading about new inventions, for example in the newspaper or on the Internet'; 'I would like to try out activities concerning design learning in class')[46]. Next to these attitudes towards S&T (teaching), teachers' self-efficacy is measured in the questionnaire, but was left out in the analyses. While the whole questionnaire was administered to the sample teachers, only items of the pleasure scale, the importance scale and the difficulty scale were retained in our analyses. In the previous large scale study of Oberon (2011), several scales that were theoretically expected were not found. This was also the case in another study by the authors (see Study 2). In the latter, results pointed in the direction of one single, underlying factor, concerning both personal as well as professional attitudes in S&T as a mixture of both

---

[42] In the questionnaire the concepts of design 'learning' and inquiry 'learning' are used. However, it is somewhat contra-intuitive, 'learning' in this questionnaire not only refer to processes in students but also to the teaching of S&T.

[43] Between these brackets, first an example of the attitude towards technology is given, and second an example of the attitude towards inquiry learning is given.

[44] Between these brackets, first an example of the attitude towards science is given, and second an example of the attitude towards design learning is given.

[45] Idem footnote 1.

[46] Idem footnote 2.

affective and cognitive elements. The behavioural dimension showed to not belong to this factor. This is also in line with the literature, as in the Theory of Planned Behavior (Ajzen & Fishbein, 1980) behavioural intention is viewed as a direct outcome of the affective and cognitive dimension of attitude, and not as a component of attitude itself. A composite score was formed by averaging the scores on the items of the different dimensions for each attitude (technology, design learning, science and inquiry learning).In our study, 26 out of 34 teachers filled in the attitude questionnaire both in the pre- and post-measurement.

Teacher style during standardised S&T-assignments

To be able to measure the initial and final style of the teacher in S&T activities, the participating primary school teachers were asked to conduct two parallel, standardised S&T-assignments with their pupils. The activities, 'Building a bridge' in the pre-measurement and 'Building a tower' in the post-measurement represented the contexts in which the quality of their classroom interactions was measured (research question 1). Both were assignments in which at most 45 minutes could be spent, where the final goal was to construct a bridge, or tower. The activities had to fulfil some requirements: only the provided paper strips (width: 5.25 cm), staplers, adhesive tape and glue sticks could be used. Moreover, the bridge had to cover a distance of 30 centimetres and the strength of the construction had to be tested by placing a small, plastic bottle, filled with 250 millilitres of water on top of it. The tower had to be made as high as possible and also had to carry a bottle with the same volume. After receiving the assignment, teachers were given approximately one week to prepare themselves to implement this activity in their classroom. Except for these requirements and the time restriction, the teachers could choose the way in which they organised and realised the activity with their students.

The classroom observations were coded with the CLASS Upper Elementary (Hamre & Pianta, 2007). With this observation instrument not only the quality of the interactions between the teacher and students but also among students is measured. The interactions with the teacher, however, are seen as the departure point. "Through these interactions, teachers act as invisible hands in the classroom, influencing children's peer behaviour both through the modelling and feedback provided during teacher-student interactions and through the ways in which the teacher uses these interactions to indirectly support and facilitate peer experiences in the classroom" (Luckner & Pianta, 2011, p. 257).

In CLASS the three domains emotional support, classroom organisation, and instructional support[47] (Hamre & Pianta, 2007; Hamre et al., 2007; Pianta et al., 2008), are each divided in smaller dimensions, which in turn are further subdivided into indicators, with the last containing some behavioural markers. In Table 1 an overview of the domains and their dimensions can be found. Firstly, the emotional support domain

[47] Next to these domains, an extra dimension, 'Student Engagement' is distinguished. This dimension is not taken into consideration in this study.

encompasses interactions that reflect the emotional climate of the classroom, mainly represented by the relationships between teachers and students and among students. Emotional support also includes teacher's awareness of and responsiveness toward students' levels of academic and social/emotional functioning and their developmental needs (NICHD ECCRN, 2002). Secondly, the classroom organisation domain includes the interactions involved in managing time, behaviour, and attention in the classroom (Hamre & Pianta, 2007; Pianta et al., 2008). Finally, the instructional support domain contains the quality of the instructional interactions, in terms of the richness of the instruction and feedback provided (Hamre & Pianta, 2005, 2007; Pianta et al., 2008). Using these indicators and markers, for each dimension a score is given on a 7-point scale. While the indicators are placed in the low, mid or high range, the behavioural markers help to situate these indicators in one of these ranges. The instructional support domain, for example, consists of a dimension 'instructional dialogue', subdivided in three indicators, namely 'cumulative content-driven exchanges', 'distributed talk' and 'facilitation strategies'. The indicator 'distributed talk' consists of four behavioural markers: student-initiated dialogues, balance of teacher and student talk, majority of students and peer dialogues.

Training is required in order to become a reliable observer[48]. The observer – the first author – followed an intensive CLASS training, obtained a certificate and passed an annual renewal test. Evidence suggests that CLASS scores, assigned by trained, certified observers, are highly reliable (Pianta et al., 2012).

[48] (http://teachstone.com/services/training/class-observation-training-programs)

Table 1. The CLASS domains (Emotional Support, Classroom Organisation and Instructional Support) with their Dimensions.

Emotional Support	**Classroom Organisation**	**Instructional Support**
Positive Climate	Behaviour Management	Instructional Learning Formats
Teacher Sensitivity	Productivity	Content Understanding
Regard for Student Perspectives	Negative Climate	Analysis and Inquiry
		Quality of Feedback
		Instructional Dialogue

Note: More detailed descriptions of the domains and their dimensions can be found in Table 1 of Hafen et al. (2015)

Two observation cycles were conducted in the activities 'Building a bridge' and 'Building a tower'. Each observation cycle consisted of 15-minute observation and 10-minute scoring. To obtain one general score for each teacher for each dimension, the available scores assigned in each observation cycle were averaged. Domain scores were calculated by averaging the dimension scores. Two teachers spent less than 45 minutes on the activities. Therefore, it was not possible to score two full cycles. Because we only used data of teachers with two cycles scored, on both cycles a missing score was given for these teachers. For the 'Building a bridge' activity a score could be calculated for 29 teachers and for the 'Building a tower' activity a score could be calculated for 27 teachers in total.

The observations of the post-measurement of the first wave were partly scored live, but most of the observations were coded on the basis of the video. The CLASS has been used before to code videotapes of classrooms (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Kane & Staiger, 2012).
To ensure intra-rater reliability[49], CLASS dimensions were evaluated again for 38 observations over all four occasions (the two parallel activities, and the two measurements during Village@School). The single measure intra-class correlation (ICC) for the three domains over all occasions varied from good to excellent (.61 for emotional support, $p = 0.000$, .84 for classroom organisation, $p = 0.000$; and .68 for instructional support, $p = 0.000$) following the criteria from Cichetti and Sparrow (1981). For the dimensions the intra-rater reliability was fair to excellent, as ICC's ranged from .42 to .90, except for the dimension 'instructional learning formats' with an ICC of .26[50]. In other studies using CLASS, ICC's (two observers)

[49] Because of the high cost of having the data double-scored by another trained and certified researcher, it was not possible to calculate inter-rater reliability. However, after having followed a training, being certified and recertified, one should be able to reliably score classroom interactions.
[50] Therefore, analyses with this dimension were excluded.

ranging from .15 to .43 are reported (Hafen et al., 2015). The ICC is a rather strict measure for the CLASS in which scoring within one point is allowed.

Teacher style during the implementation of Village@School

Teacher style, operationalised in quality of classroom interactions, was also measured in the intervention, Village@School. In contrast to the activities in the pre- and post-measurement, which were much shorter, Village@School had to be implemented for a minimum of 20 weeks and teachers were supported during this implementation.

Two Village@School sessions were coded live, but these observations were also video-recorded in order to have a backup (see further). Lessons were recorded with one camera with a wireless microphone attached to the teachers. This made it possible to capture teachers' interactions with pupils, which is necessary in group work settings where pupils are talking with each other and interactions with the teacher are not always clearly audible. The camera was positioned sideways in the classroom, in order to have a global view of the class. By that, not only interactions with the teacher, but also interactions among students could be observed. No focus was specifically given to individual pupils, rather, interactions among the whole class group were taken into consideration.

Two Village@School sessions (approximately one hour observation) were coded live, resulting in two observation cycles for each session.[51] It is important to note that the video-scores did not differ significantly from the live scores for the domains. A paired samples t-test showed no significant p-values for the three CLASS-domains ($p = .427$ for emotional support, $p = .457$ for classroom organisation and $p = .694$ for instructional support).

[51] However, because of practical reasons (e.g. teachers stopped earlier with the session) some cycles couldn't be fully conducted because there were for example only 5 minutes of the session left after scoring the first 15 minutes of the observation. Because the video-recording allows to stop watching the video after a 15-minutes observation and starting again at that time after scoring the first observation, it was possible to score more cycles via the video or select another cycle (not mainly containing the clean-up of teachers and students).

Data analysis

Both teachers' attitudes and their classroom interactions are situated at the class level (research question 1). However, because these variables may also be partly determined by the school in which teachers operate, data were analysed by means of multilevel modelling techniques (Goldstein, 1995), making use of the software program MLwiN (Rashbash et al., 2005). Three levels of information are involved in this study: time (level 1), nested within classes (level 2) and nested within schools (level 3).

First, baseline models were tested in the prediction of every CLASS domain, measured during Village@School, to partition the variance for each of the outcomes between school, class and time level. The last time level was included because interactions were coded in two Village@School sessions. When school and/or class level showed to be significant, the four attitudes - respectively science, technology, inquiry learning, design learning – were separately were added as predictors (see Models A) and then all together (see Models B). the same was done for the CLASS domains as evaluated in the pre-measurement. Finally, all variables of the teachers' competence profile (initial attitudes and CLASS domains) were added all together (Models C).

Second, a paired samples t-test was conducted to test for a significant change in the participating teachers' attitudes towards S&T (teaching) and their teacher style between the pre- and post-measurement (research question 2).

Third, baseline models were tested for the CLASS domains and dimensions as measured over all occasions (pre, two in-between measurements, post) in order for partitioning the variance for each of the outcomes between school, class and time level. When there was a significant level of variance on the time level, a predictor time – centred around the grand mean – was created and tested in the model (Model A). In case of a significant time predictor and a significant amount of variance at the class level, we allowed the effect of time to vary across classes in order to determine whether classes differ in their eventual growth concerning a particular CLASS domain or dimension.

Results

Relation between teachers' competence profile and teacher style during Village@School (Research question 1)

In Tables 2, 3 and 4 descriptive statistics are presented.

Table 2. Means, Standard Deviations, Range, Skewness and Kurtosis of Study Variables (CLASS-domains during Village@School as outcome)

	M	SD	Range	Skewness (SE)	Kurtosis (SE)
1. Emotional Support (Village@School)_two_measurement_occasions	5.00	.48	2.00	-1.17(.32)	.90 (.63)
2. Classroom Organisation (Village@School)_two_measurement_occasions	5.99	.32	1.33	-1.11 (.32)	.93(.63)
3. Instructional Support (Village@School)_two_measurement_occasions	4.56	.56	3.00	-.51(.32)	.87(.63)
4. Emotional Support (Village@School)_measurement_occasion_1	5.13	.42	1.84	-1.26(.44)	2.10(.86)
5. Emotional Support (Village@School)_measurement_occasion_2	4.88	.51	1.83	-1.08(.45)	.14(.87)
6. Classroom_Organisation (Village@School)_measurement_occasion_1	6.03	.22	.83	-.47(.44)	-.02(.86)
7. Classroom Organisation (Village@School)_measurement_occasion_2	5.94	.41	1.33	-.90(.45)	-.30(.87)
8. Instructional Support (Village@School)_measurement_occasion_1	4.47	.58	2.60	-.81(.44)	1.06(.86)
9. Instructional Support (Village@School)_measurement_occasion_2	4.66	.53	2.20	-.05(.45)	.21(.87)
10. Emotional Support_pre	4.80	.31	1.17	.40(.43)	-.65(.85)
11. Classroom_Organisation_pre	5.87	.40	1.66	-1.64(.43)	2.75(.85)
12. Instructional_Support_pre	4.55	.47	2.30	.63(.43)	1.19(.85)
13. Attitude T_pre	3.79	.41	1.93	-.72(.43)	1.06(.85)
14. Attitude DL_pre	3.31	.45	2.00	-.00(.43)	-.07(.85)
15. Attitude S_pre	3.46	.45	1.57	-.28(.43)	-1.02(.85)
16. Attitude_IL_pre	3.99	.33	1.19	.44(.43)	-.56(.85)

Note: T = Technology; DL = Design Learning; S = Science; IL = Inquiry Learning

Table 3. Bivariate correlations between Study Variables (CLASS-domains during Village@School_two_measurement_occasions as outcome)

	1	2	3	4	5	6	7	8	9	10
1. Emotional Support (Village@School)_two_measurement_occasions	1									
2. Classroom Organisation (Village@School)_two_measurement_occasions	.54**	1								
3. Instructional Support (Village@School)_two_measurement_occasions	.32*	.43**	1							
4. Emotional Support_pre	.18	-.02	-.04	1						
5. Classroom_Organisation_pre	-.09	.05	-.06	.45**	1					
6. Instructional_Support_pre	.03	.07	.11	.36**	.60**	1				
7. Attitude_T_pre	-.10	.04	-.04	.20	.10	.10	1			
8. Attitude_DL_pre	-.03	.05	-.02	.31	.12	.17	.70**	1		
9. Attitude_S_pre	.09	.08	.18	.23	-.04	.03	.69**	.58**	1	
10. Attitude_IL_pre	-.14	-.28*	-.12	.11	-.51**	-.24	.23	.34	.21	1

Note: T = Technology; DL = Design Learning; S = Science; IL = Inquiry Learning

*p < .05, **p < .01

Table 4. Bivariate correlations between Study Variables (CLASS-domains during Village@School, measurement occasion 1 and 2 as outcome)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Emotional Support (Village@School)_ measurement_occasion_1	1												
2. Emotional Support (Village@School)_measurement_occasion_2	.62**	1											
3. Classroom_Organisation (Village@School)_measurement_occasion_1	.24	.18	1										
4. Classroom Organisation (Village@School)_measurement_occasion_2	.35	.68**	.27	1									
5. Instructional Support (Village@School)_measurement_occasion_1	.13	.18	.35	.39*	1								
6. Instructional Support (Village@School)_measurement_occasion_2	.20	.63**	.10	.59**	.33	1							
7. Emotional Support_pre	.22	.15	-.04	-.02	.04	-.12	1						
8. Classroom_Organisation_pre	-.25	.03	-.17	.18	-.15	.06	.45**	1					
9. Instructional_Support_pre	-.01	.06	-.01	.12	.20	.01	.36**	.60**	1				
10. Attitude_T_pre	-.18	-.05	.16	-.02	.06	-.15	.20	.10	.10	1			
11. Attitude_DL_pre	-.04	-.04	.35	-.13	.06	-.09	.31	.12	.17	.70**	1		
12. Attitude_S_pre	-.09	.25	.21	.00	.23	.13	.23	-.04	.03	.69**	.58**	1	
13. Attitude_IL_pre	.04	-.30	.15	-.53**	.03	-.30	.11	-.51**	-.24	.23	.34	.21	1

Note: T = Technology; DL = Design Learning; S = Science; IL = Inquiry Learning

*p < .05, **p < .01

Estimation of the random part of the baseline models showed that there is significant variance at the school level for the emotional support domain as measured during two sessions of the implementation of Village@School (see Table 5). As explained in the analysis section, the attitudes and CLASS domains, separately and together, were added to the model in order to determine whether they explain this school level variance (see Tables 5, 6 and 7). For the other two CLASS-domains no significant variance at the school and/or class level was found (see Table 8).

Table 5: Predicting the CLASS domain Emotional Support during Village@School by Attitudes Towards Technology (T), Science (S), Design Learning (DL) and Inquiry Learning (IL)

	Emotional Support (Village@School) (n=55)[a]													
	Baseline Model		Model A1[c]			Model A2			Model A3			Model A4		
	B	SE	B	SE	Δ R2	B	SE	Δ R2	B	SE	Δ R2	B	SE	Δ R2
Random parameters					.00			.00			.00			.00
School level														
σ_{cons}^2	.11*	.05	.11*	.05		.11*	.05		.11*	.05		.11*	.05	
Class level														
σ_{cons}^2														
Time level														
σ_{cons}^2	.11***	.03	.11***	.03		.11***	.03		.11***	.03		.11***	.03	
Deviance	65.661		65.294			65.642			65.283			64.866		
					β			β			β			β
Fixed parameters														
Intercept	5.01***	.08	5.00***	.08		5.01***	.08		5.01***	.08		5.01***	.08	
Attitude T			-.12	.19	-.10									
Attitude DL						-.02	.17	-.02						
Attitude S									.11	.18	.10			
Attitude IL												-.21	.23	-.14

[a] In the prediction of Emotional Support during Village@School, data were available on all predictor variables for 29 teachers, yielding (29 x 2 occasions) = 58 data points of information. Over the 2 occasions, 3 pieces of information were missing on Emotional Support during Village@School (leading to n =55).

[b]The baseline model did not include any predictors.

[c]In Models A (A1 to A4), each attitude was added to the model separately.

Table 6: Predicting the CLASS domain Emotional Support during Village@School by CLASS Domain Scores in the Pre-measurement

	Emotional Support (Village@School) (n=55)[a]											
	Baseline Model			Model A5[c]			Model A6			Model A7		
	B	SE		B	SE	Δ R^2	B	SE	Δ R^2	B	SE	Δ R^2
Random parameters						.00			.00			.00
School level												
σ_{cons}^2	.11*	.05		.11*	.05		.11*	.05		.11*	.05	
Class level												
σ_{cons}^2												
Time level												
σ_{cons}^2	.11***	.03		.11***	.03		.11***	.03		.11***	.03	
Deviance	65.661			64.344			65.413			65.639		
			β			β			β			β
Fixed parameters												
Intercept	5.01***	.08		3.64**	1.18		5.58***	1.14		4.89***	.76	
Emotional_Support_pre				.29	.25	.19						
Classroom_Organisation_pre							-.10	.20	-.08			
Instructional_Support_pre										.03	.17	.03

[a] In the prediction of Emotional Support during Village@School, data were available on all predictor variables for 29 teachers, yielding (29 x 2 occasions) = 58 data points of information. Over the 2 occasions, 3 pieces of information were missing on Emotional Support during Village@School (leading to n =55).

[b]The baseline model did not include any predictors.

[c]In Models A (A5 to A8), each CLASS Domain score (pre-measurement) was added to the model separately.

Table 7: Predicting the CLASS domain Emotional Support during Village@School by Attitudes Towards Technology (T), Science (S), Design Learning (DL) and Inquiry Learning (IL) and CLASS Domain Scores in the Pre-measurement

	Emotional Support (Village@School) (n=55)[a]											
	Baseline Model			Model B1[c]			Model B2			Model C		
	B	SE		B	SE	Δ R2	B	SE	Δ R2	B	SE	Δ R2
Random parameters						.02			.01			.05
School level												
σ_{cons}^2	.11*	.05		.09*	.04		.10*	.04		.06	.03	
Class level												
σ_{cons}^2												
Time level												
σ_{cons}^2	.11***	.03		.11***	.03		.11***	.03		.11***	.03	
Deviance	65.661			62.180			62.834			54.350		
			β			β			β			β
Fixed parameters												
Intercept	5.01***	.08		5.00***	.07		5.01***	.07		5.00***	.06	
Attitude T				-.42	.30	-.36				-.32	.26	-.27
Attitude DL				.13	.26	.12				.23	.24	.23
Attitude S				.33	.23	.31				.14	.21	.13
Attitude IL				-.25	.24	-.17				-.78**	.29	-.53**
Emotional_Support_pre							.41	.27	.26	.69*	.27	.44*
Classroom_Organisation_pre							-.30	.25	-.25	-.74**	.28	-.61**
Instructional_Support_pre							.08	.20	.08	.09	.17	.09

[a] In the prediction of Emotional Support during Village@School, data were available on all predictor variables for 29 teachers, yielding (29 x 2 occasions) = 58 data points of information. Over the 2 occasions, 3 pieces of information were missing on Emotional Support during Village@School (leading to n =55).

[b]The baseline model did not include any predictors.

[c]In Models B (B1 and B2), all four attitudes respectively all three CLASS-domains (pre) were added to the model.

[d]In Model C, all four attitudes and all three CLASS-domains (pre) were added to the model.

Table 8: No variance at the school and class level for the CLASS domains Classroom Organisation and Instructional Support during Village@School

	Classroom Organisation (Village@School) (n=55)[a] Baseline Model[b]		Instructional Support (Village@School) (n=55) Baseline Model	
	B	SE	B	SE
Random parameters				
School level				
σ_{cons}^2	.02	.02	.09	.06
Class level				
σ_{cons}^2	.00	.00	.00	.00
Time level				
σ_{cons}^2	.08***	.02	.22***	.06
Deviance	29.741		88.595	
Fixed parameters				
Intercept	5.99***	.05	4.56***	.08

[a] In the prediction of Classroom Organisation and Instructional Support during Village@School, data were available on all predictor variables for 29 teachers, yielding (29 x 2 occasions) = 58 data points of information. Over the 2 occasions, 3 pieces of information were missing on Emotional Support during Village@School (leading to n =55).

[b]The baseline model did not include any predictors.

Concerning Models A, in which each predictor was added separately (Tables 5 and 6), and B, in which the attitudes respectively the CLASS domains were added all together (Table 7), no effects were found on the quality of the emotional support provided during Village@School. On the contrary, taking all attitudes and CLASS-domains into account in Model C shows a positive effect of the emotional support as provided in the S&T-assignment but a negative effect of classroom organisation, as well as a negative effect of the attitude towards inquiry learning (see Table 7). Schools characterized by an initially higher emotional support also have a higher emotional support during the implementation of Village@School, when controlling for the other CLASS-domains and the attitudes. Under the same conditions, schools with a higher classroom organisation and a better attitude towards inquiry learning of their teachers are related to a lower emotional support when implementing Village@School. Some amount of variance is explained by the emotional support provided during Village@School (see Table 4). When the attitudes and CLASS-domains were added simultaneously to the baseline model as predictors, the deviance of the total model did not reduce significantly (X^2 (7) = 11.31).

As some of these results are somewhat unexpected, further analyses were conducted in order to examine the presence of outliers that may bias the results. These analyses show that the amount of outliers is limited. For the attitude predictor variables, no outliers were detected for the attitude towards science and the attitude towards inquiry learning. Only for the attitude towards technology and the attitude towards design learning one outlier was found (at the low end for the attitude towards technology with a value of 2.60; at the high end for the attitude towards design learning with a value of 4.33). Concerning the CLASS-domain variables in the pre-measurement, for two domains - Classroom Organisation and Instructional Support - outliers were detected (two extreme values at the low end for classroom organisation (with values of 4.67 and 4.83) and one at the high end for instructional support (with a value of 5.80)). For these domains just as many outliers were found for the outcome variables varying over the two Village@School occasions (a score of 5.00 and 5.17 for Classroom Organisation and a score of 2.80 for Instructional Support). However, for the outcome variable Emotional Support four outliers were detected at the low end. Two of these scores, for each of the two measurement occasions, are from the same class (with scores of 3.67 and 3.83).

Change in teachers' competence profile (research question 2)

Change in teachers' attitudes towards S&T (teaching)

In Tables 9 and 10 descriptive statistics and correlations are presented.

Table 9. Means, Standard Deviations, Range, Skewness and Kurtosis of Study Variables (Teacher Attitudes (pre- and post-measurement))

	M	SD	Range	Skewness (SE)	Kurtosis (SE)
1. Attitude_T_pre	3.80	.41	1.93	-.72(.42)	1.06(.82)
2. Attitude_DL_pre	3.33	.45	2.00	-.07(.42)	-.08(.82)
3. Attitude_S_pre	3.45	.44	1.57	-.21(.42)	-.96(.82)
4. Attitude_IL_pre	4.00	.33	1.19	.41(.42)	-.57(.82)
5. Attitude_T_post	3.75	.38	1.33	-.46(.44)	-.74(.86)
6. Attitude_DL_post	3.27	.41	1.83	.33(.44)	.79(.86)
7. Attitude_S_post	3.52	.48	2.00	.17(.44)	-.24(.86)
8. Attitude_IL_post	3.85	.38	1.64	.17(.44)	-.24(.86)

Note: T = Technology; DL = Design Learning; S = Science; IL = Inquiry Learning

Table 10. Bivariate correlations between Study Variables (Teacher Attitudes (pre- and post-measurement)

	1	2	3	4	5	6	7	8
1. Attitude_T_pre	1							
2. Atttitude_DL_pre	.65**	1						
3. Attitude_S_pre	.68**	.53**	1					
4. Attitude_IL_pre	.19	.36*	.19	1				
5. Attitude_T_post	.56**	.59**	.33	.03	1			
6. Attitude_DL_post	.32	.50**	.37	.26	.40*	1		
7. Attitude_W_post	.70**	.63**	.58**	.39	.59**	.44*	1	
8. Attitude_IL_post	.36	.40*	.26	.49*	.40*	.52**	.71**	1

Note: T = Technology; DL = Design Learning; S = Science; IL = Inquiry Learning
*p < .05, **p < .01

Teachers did not make any significant gains in their attitudes towards S&T (teaching) after the course of the Village@School project compared to before the project, as evidenced by their responses on the attitude questionnaire (see Table 11).

Table 11. Change in Teachers' Attitudes towards S&T (teaching)

Attitudes	**Pre-measurement mean**	**Post-measurement mean**	**SD**	**Difference**	**Cohen's d**	**T-test**		
						t	***df***	***p***
Technology	3.72	3.75	.07	.03	.08		25	.70
Science	3.42	3.56	.08	.14	.31		25	.12
Design Learning	3.33	3.28	.09	-.05	-.10			
Inquiry Learning	3.98	3.88	.07	-.10	-.28		25	.16

Note: For the pre-measurement, data was missing for 3 of the 34 teachers; for the post-measurement data were missing for 6 of the 34 teachers.

Change in teacher style

In Tables 12 and 13 descriptive statistics and correlations are presented with regard to teacher style in the pre- and post-measurement.

Table 12. Means, Standard Deviations, Range, Skewness and Kurtosis of Study Variables (CLASS-dimensions (pre- and post-measurement))

	M	SD	Range	Skewness (SE)	Kurtosis (SE)
1. Positive Climate_pre	4.95	.51	2.00	.44(.43)	-.48(.85)
2. Teacher Sensitivity_pre	5.12	.55	2.00	.01(.43)	-.01(.85)
3. Regard_For_Student_Perspectives_pre	4.33	.36	1.00	.64(.43)	-.76(.85)
4. Behavior_Managament_pre	5.59	.71	3.00	-2.11(.43)	5.22(.85)
5. Productivity_pre	5.22	.59	2.50	-1.12(.43)	1.52(.85)
6. Negative_Climate_Rev_pre	6.79	.47	2.00	-2.58(.43)	6.91(.85)
7. Content_Understanding_pre	4.10	.89	3.50	-.42(.43)	-.19(.85)
8. Analysis_and_Inquiry_pre	4.53	.74	2.50	.36(.43)	-.82(.85)
9. Quality_of_Feedback_pre	4.17	.57	2.00	-.28(.43)	-.42(.85)
10. Instructional Dialogue_pre	4.69	.69	2.50	.40(.43)	-.30(.85)
11. Positive_Climate_post	5.02	.63	3.00	-1.34(.45)	3.22(.87)
12. Teacher_Sensitivity_post	5.15	.53	2.50	-1.47(.45)	2.71(.87)
13. Regard_for_Student_Perspectives_post	4.83	.39	1.50	-.34(.45)	.03(.87)
14. Behavior_Management_post	5.56	.64	2.50	-1.65(.45)	2.75(.87)
15. Productivity_post	5.31	.46	2.00	-.73(.45)	1.27(.87)
16. Negative_Climate_Rev_post	6.33	.71	2.50	-.94(.45)	.19(.87)
17. Content_Understanding_post	4.33	.77	3.00	-.61(.45)	.71(.87)
18. Analysis_and_Inquiry_post	4.69	.67	3.00	-1.27(.45)	3.03(.87)
19. Quality_of_Feedback_post	4.46	.82	3.00	-.72(.45)	-.26(.87)
20. Instructional_Dialogue_post	4.76	.54	3.00	-.05(.45)	.42(.87)

Table 13. Bivariate correlations between Study Variables (CLASS-dimensions (pre- and post-measurement)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1.	Positive Climate_pre	1																			
2.	Teacher Sensitivity_pre	.19	1																		
3.	Regard_For_Student_Perspectives_pre	.10	.11	1																	
4.	Behavior_Management_pre	.04	.32	.31	1																
5.	Productivity_pre	.34	.38*	.15	.27	1															
6.	Negative_Climate_Rev_pre	.03	.24	-.22	-.16	.40*	1														
7.	Content_Understanding_pre	.19	.14	.20	.31	.46*	.31	1													
8.	Analysis_and_Inquiry_pre	-.14	.32	.29	.23	.29	.38*	.39*	1												
9.	Quality_of_Feedback_pre	.19	.22	.11	.07	.29	.30	.39*	.45*	1											
10.	Instructional_ Dialogue_pre	-.05	.18	.57**	.13	.23	.26	.07	.27	.39*	1										
11.	Positive_Climate_post	.10	.25	-.24	.06	-.06	.15	.08	.23	.31	-.01	1									
12.	Teacher_Sensitivity_post	.18	.45*	.03	.12	.46*	.47*	.39	.34	.26	.17	.42*	1								
13.	Regard_for_Student_Perspectives_post	-.04	-.16	-.04	-.03	-.21	-.44*	-.52**	-.16	-.12	-.24	-.18	-.57**	1							
14.	Behavior_Management_post	-.24	.31	-.20	.06	.12	.38	-.04	.25	.31	.34	.45*	.17	-.34	1						
15.	Productivity_post	.12	.22	.33	.31	.10	.01	.24	.20	-.01	.19	.18	.43*	-.34	-.06	1					
16.	Negative_Climate_Rev_post	.16	.03	-.13	.14	.09	.25	.36	.15	.36	-.03	.72**	.35	-.38*	.38*	.08	1				
17.	Content_Understanding_post	.33	.28	-.18	.00	.37	.45*	.28	.41*	.55**	-.13	.09	.32	.13	-.04	.21	.02	1			
18.	Analysis_and_Inquiry_post	.12	-.09	.16	-.05	.04	.33	.34	.37	.39	.02	.27	.22	.05	.04	.18	.21	.47*	1		
19.	Quality_of_Feedback_post	.11	.09	-.15	-.13	-.21	.27	.10	.50*	.52**	-.09	.47*	.12	.13	.19	.16	.19	.58**	.61**	1	
20.	Instructional_Dialogue_post	-.02	-.24	-.09	-.09	.01	.30	.17	.23	.56**	.02	.13	.16	-.02	.26	-.03	.37	.50**	.55**	.50**	1

*p < .05, **p < .01

Table 14 shows that teachers grew significantly in three dimensions after the implementation of the project-based learning environment Village@School in comparison to before the project. First, teachers became more open to pupils' ideas and realities and gave them more responsibility (Regard for Student Perspectives). Second, the quality of feedback provided in the classroom improved, meaning that teachers and/or pupils posed more questions to each other, and they engaged in persistent feedback loops in order to gain deeper insight into the particular content they were confronted with (Quality of Feedback). Third, Negative Climate[52] (reversed) shows a decrease after the project, meaning that over the course of the trajectory teachers and/or pupils showed more irritation and exclusion in interactions with each other.

Table 14. Change in Teacher Style when Teaching S&T

CLASS dimensions	**Pre-measurement mean**	**Post-measurement mean**	**SD**	**Difference**	**Cohen's d**	**t-test[a]**		
						t	*df*	*p*
Positive Climate	4.94	5.02	.82	.08	.10	.50	23	.622
Teacher Sensitivity	5.06	5.15	.56	.08	.15	.72	23	.477
Regard for Student Perspectives	4.31	4.85	.57	.54	.95	4.66***	23	.000
Behaviour Management	5.52	5.52	.98	.00	.00	.00	23	1.000
Productivity	5.20	5.31	.77	.10	.14	.67	23	.512
Negative Climate (reversed)	6.79	6.31	.77	-.48	-.62	-3.04**	23	.006
Content Understanding	4.06	4.38	1.06	.31	.29	1.44	23	.163
Analysis and Inquiry	4.54	4.67	.82	.13	.15	.74	23	.47
Quality of Feedback	4.10	4.42	.73	.31	.43	2.08*	23	.048
Instructional Dialogue	4.73	4.65	.87	.08	.10	.47	23	.643

Note: For the pre-measurement data were missing for 5 of the 34 teachers; for the post-measurement data was missing for 7 of the 34 teachers.

[a] Paired-Samples t-test

*p < 0.05, **p < 0.01; *** p < 0.001

In Tables 15, 16, 17, 18, 19 and 20 descriptive statistics and correlations for the CLASS-domains and dimensions over all measurement occasions are presented.

[52] Negative Climate in the CLASS is different from Positive Climate, as measured in the CLASS. In Negative Climate, the irritation, frustration, sarcasm etc. in the comments of the teacher and other pupils is evaluated, whereas in Positive Climate the quality of the relationships, respect, giving positive comments etc. are scored (Pianta et al., 2012).

Table 15. Means, Standard Deviations, Range, Skewness and Kurtosis of Study Variables (CLASS-domains and dimensions (all measurement occasions, together and separately))

	M	SD	Range	Skewness (SE)	Kurtosis (SE)
1. Positive_Climate_all_measurement_occasions	4.90	.70	3.00	-.72(.23)	.71(.45)
2. Teacher_ Sensitivity_all_measurement_occasions	5.15	.59	3.00	-.62(.23)	.78(.45)
3. Regard_For_Student_Perspectives_pre_all_measurement_occasions	4.81	.55	3.00	-.03(.23)	.52(.45)
4. Behavior_Management_pre_all_measurement_occasions	5.67	.55	3.00	-2.21(.23)	6.00(.45)
5. Productivity_all_measurement_occasions	5.45	.51	2.50	-1.07(.23)	1.41(.45)
6. Negative_Climate_Rev_all_measurement_occasions	6.56	.67	3.50	-2.02(.23)	4.68(.45)
7. Content_Understanding _all_measurement_occasions	4.06	.83	4.00	-.43(.23)	-.14(.45)
8. Analysis_and_Inquiry_all_measurement_occasions	4.50	.78	4.00	-.40(.23)	.10(.45)
9. Quality_of_Feedback_all_measurement_occasions	4.32	.77	3.50	-.10(.23)	-.44(.45)
10. Instructional Dialogue_all_measurement_occasions	4.75	.64	3.50	-.17(.23)	.60(.45)
11. Emotional_Support_all_measurement_occasions	4.95	.40	2.00	-.81(.23)	.70(.45)
12. Classroom_Organisation_all_measurement_occasions	5.89	.38	1.83	-1.28(.23)	1.98(.45)
13. Instructional_Support_all_measurement_occasions	4.58	.52	3.00	-.42(.23)	.77(.45)
14. Positive_ Climate_pre	4.95	.51	2.00	.44(.43)	-.48(.85)
15. Teacher_Sensitivity_pre	5.12	.55	2.00	.01(.43)	-.01(.85)
16. Regard_For_Student_Perspectives_pre	4.33	.36	1.00	.64(.43)	-.76(.85)
17. Behavior_Management_pre	5.59	.71	3.00	-2.11(.43)	5.22(.85)
18. Productivity_pre	5.22	.59	2.50	-1.12(.43)	1.52(.85)
19. Negative_Climate_Rev_pre	6.79	.47	2.00	-2.59(.43)	6.91 (.85)
20. Content_Understanding_pre	4.10	.89	3.50	-.42(.43)	-.19(.85)
21. Analysis_and_Inquiry_pre	4.53	.74	2.50	.36(.43)	-.82(.85)
22. Quality_of_Feedback_pre	4.17	.57	2.00	-.28(.43)	-.42(.85)
23. Instructional Dialogue_pre	4.69	.69	2.50	.40(.43)	-.30(.85)
24. Emotional_Support_pre	4.80	.31	1.17	.40(.43)	-.65(.85)
25. Classroom_Organisation_pre	5.87	.40	1.66	-1.64(.43)	2.75(.85)
26. Instructional_Support_pre	4.55	.47	2.30	.63(.43)	1.19(.85)
27. Positive_Climate (Village@School)_measurement_ occasion1	4.97	.78	3.00	-.53(.43)	.13(.83)
28. Teacher_Sensitivity (Village@School)_measurement_occasion 1	5.27	.64	2.00	-.29(.43)	-.93(.83)
29. Regard_For_Student_Perspectives (Village@School)_measurement_occasion 1	5.12	.57	2.00	-.04(.43)	-.47(.83)

	M	SD	Range	Skewness (SE)	Kurtosis (SE)
30. Behavior_Management (Village@School)_measurement_occasion 1	5.75	.34	1.00	-1.05(.43)	-.03(.83)
31. Productivity (Village@School)_measurement_occasion 1	5.58	.40	1.50	-.76(.43)	.42(.83)
32. Negative_Climate_Rev (Village@School)_measurement_occasion 1	6.67	.38	1.00	-.66(.43)	-.91(.83)
33. Content_Understanding (Village@School)_measurement_occasion_1	3.78	.91	3.00	-.41(.43)	-.34(.83)
34. Analysis_and_Inquiry (Village@School)_measurement_occasion_1	4.32	.95	4.00	-.49(.43)	-.32(.83)
35. Quality_of_Feedback (Village@School)_measurement_occasion_1	4.18	.78	3.00	-.03(.43)	-.55(.83)
36. Instructional Dialogue (Village@School)_measurement_occasion_1	4.77	.72	3.50	-.68(.43)	2.42(.83)
37. Emotional_Support (Village@School)_measurement_occasion_1	5.12	.40	1.84	-1.23(.43)	2.25(.83)
38. Classroom_Organisation (Village@School)_measurement_occasion_1	6.00	.24	.83	-.45(.43)	-.38(.83)
39. Instsructional_Support (Village@School)_measurement_occasion_1	4.45	.57	2.60	-.75(.43)	1.06(.83)
40. Positive Climate (Village@School)_measurement_ occasion2	4.68	.80	3.00	-.70(.43)	-.13(.83)
41. Teacher_Sensitivity (Village@School)_measurement_occasion 2	5.05	.65	3.00	-1.01(.43)	2.05(.83)
42. Regard_For_Student_Perspectives (Village@School)_measurement_occasion 2	4.93	.50	3.00	-1.45(.43)	7.73(.83)
43. Behavior_Management (Village@School)_measurement_occasion 2	5.77	.45	2.00	-2.54(.43)	7.57(.83)
44. Productivity (Village@School)_measurement_occasion 2	5.65	.48	1.50	-1.43(.43)	1.34(.83)
45. Negative_Climate_Rev (Village@School)_measurement_occasion 2	6.43	.91	3.50	-1.93(.43)	3.32(.85)
46. Content_Understanding (Village@School)_measurement_occasion_2	4.05	.69	2.00	-.02(.43)	-1.00(.83)
47. Analysis_and_Inquiry (Village@School)_measurement_occasion_2	4.47	.72	3.00	.20(.43)	-.44(.83)
48. Quality_of_Feedback (Village@School)_measurement_occasion_2	4.48	.86	3.00	.01(.43)	-.49(.83)
49. Instructional_Dialogue (Village@School)_measurement_occasion_2	4.77	.64	2.50	-.19(.43)	-.19(.83)
50. Emotional_Support (Village@School)_measurement_occasion_2	4.89	.49	1.83	-1.19(.43)	.53(.83)
51. Classroom_Organisation (Village@School)_measurement_occasion_2	5.94	.41	1.33	-.82(.43)	-.48(.85)
52. Instructional_Support (Village@School)_measurement_occasion_2	4.63	.53	2.20	-.09(.43)	.11(.83)
53. Positive_Climate_post	5.02	.63	3.00	-1.34(.45)	3.22(.87)
54. Teacher_Sensitivity_post	5.15	.53	2.50	-1.47(.45)	2.71(.87)
55. Regard_for_Student_Perspectives_post	4.83	.39	1.50	-.34(.45)	.03(.87)
56. Behavior_Management_post	5.56	.64	2.50	-1.65(.45)	2.75(.87)
57. Productivity_post	5.31	.46	2.00	-.73(.45)	1.27(.87)
58. Negative_Climate_Rev_post	6.33	.71	2.50	-.94(.45)	.19(.87)
59. Content_Understanding_post	4.33	.77	3.00	-.61(.45)	.71(.87)
60. Analysis_and_Inquiry_post	4.69	.67	3.00	-1.27(.45)	3.03(.87)
61. Quality_of_Feedback_post	4.46	.82	3.00	-.72(.45)	-.26(.87)
62. Instructional_Dialogue_post	4.76	.54	2.50	-.05(.45)	.42(.87)
63. Emotional_Support_post	5.00	.30	1.17	-.68(.45)	.58(.87)
64. Classroom_Organisation_post	5.73	.41	1.83	-1.39(.45)	2.64(.87)
65. Instructional_Support_post	4.69	.49	2.20	-1.26(.45)	2.07(.87)

Table 16. Bivariate correlations between Study Variables (CLASS-domains and –dimensions, all measurement occasions)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Positive Climate_all_measurement_occasions	1												
2. Teacher Sensitivity_all_measurement_occasions	.33**	1											
3. Regard_For_Student_Perspectives_pre_all_measurement_occasions	.12	-.11	1										
4. Behavior_Management_pre_all_measurement_occasions	.16	.14	.13	1									
5. Productivity_all_measurement_occasions	.11	.22*	.16	.19*	1								
6. Negative_Climate_Rev_all_measurement_occasions	.47**	.35**	-.08	.11	.11	1							
7. Content_Understanding _all_measurement_occasions	.08	.06	-.01	.13	.18	.05	1						
8. Analysis_and_Inquiry_all_measurement_occasions	.06	.18	.09	.17	.16	.13	.46**	1					
9. Quality_of_Feedback_all_measurement_occasions	.42**	.18	.18	.20*	.15	.25**	.52**	.49**	1				
10. Instructional Dialogue_all_measurement_occasions	.19*	.16	.37**	.24**	.17	.33**	.28**	.48**	.51**	1			
11. Emotional_Support_all_measurement_occasions	.80**	.64**	.47**	.22*	.25**	.41**	.07	.16	.41**	.35**	1		
12. Classroom_Organisation_all_measurement_occasions	.40**	.39**	.08	.64**	.61**	.70**	.17	.23*	.31**	.39**	.46**	1	
13. Instructional_Support_all_measurement_occasions	.25**	.24*	.18	.26**	.25**	.25**	.75**	.79**	.80**	.69**	.34**	.39**	1

*p < .05, **p < .01

Table 17. Bivariate correlations between Study Variables (CLASS-variables of each measurement occasion with CLASS-variables in premeasurement)

		1	2	3	4	5	6	7	8	9	10	11	12	13
1.	Positive Climate_pre	1												
2.	Teacher Sensitivity_pre	.19	1											
3.	Regard_For_Student_Perspectives_pre	.10	.11	1										
4.	Behavior_Management_pre	.04	.32	.31	1									
5.	Productivity_pre	.34	.38*	.15	.27	1								
6.	Negative_Climate_Rev_pre	.03	.24	-.22	-.16	.40*	1							
7.	Content_Understanding_pre	.19	.14	.20	.31	.46*	.31	1						
8.	Analysis_and_Inquiry_pre	-.14	.32	.29	.23	.29	.38*	.39*	1					
9.	Quality_of_Feedback_pre	.19	.22	.11	.07	.23	.30	.39*	.45*	1				
10.	Instructional Dialogue_pre	-.05	.18	.57**	.13	.29	.26	.07	.27	.40*	1			
11.	Emotional_Support_pre	.69**	.73**	.51**	.33	.47*	.07	.26	.23	.27	.30	1		
12.	Classroom_Organisation_pre	.20	.48**	.17	.67**	.82**	.50**	.54**	.43*	.28	.33	.45*	1	
13.	Instructional_Support_pre	.05	.33	.37*	.29	.50**	.45*	.72**	.77**	.72**	.54**	.36	.60**	1
14.	Positive Climate (Village@School)_measurement_ occasion1	.31	-.10	.15	-.06	.09	-.13	-.13	-.17	-.31	.06	.17	-.05	-.15
15.	Teacher Sensitivity (Village@School)_measurement_occasion 1	-.01	.38*	.14	-.10	-.06	-.03	-.25	.10	.20	.31	.27	-.11	.09
16.	Regard_For_Student_Perspectives (Village@School)_measurement_occasion 1	.02	-.22	.09	-.31	-.29	-.05	.09	.11	-.03	.07	-.08	-.34	.09
17.	Behavior_Management (Village@School)_measurement_occasion 1	-.07	-.04	-.12	-.16	.22	.36	.16	.44*	.35	.18	-.11	.16	.41*
18.	Productivity (Village@School)_measurement_occasion 1	.28	.14	.24	-.24	-.17	-.15	-.20	.14	.02	-.02	.33	-.29	-.05
19.	Negative_Climate_Rev (Village@School)_measurement_occasion 1	-.04	-.40*	-.14	-.13	-.09	-.12	-.14	-.34	-.26	-.29	-.31	-.17	-.35
20.	Content_Understanding (Village@School)_measurement_occasion_1	.11	-.11	-.08	-.21	.00	.19	.13	.32	.12	-.24	-.03	-.05	.09
21.	Analysis_and_Inquiry (Village@School)_measurement_occasion_1	-.11	-.02	.38*	.00	-.23	.03	.09	.51**	.32	.05	.08	-.10	.27
22.	Quality_of_Feedback (Village@School)_measurement_occasion_1	.09	-.14	-.01	-.01	.07	-.02	-.05	.39*	.21	-.08	-.03	.02	.13
23.	Instructional Dialogue (Village@School)_measurement_occasion_1	.12	-.28	.20	-.33	-.20	-.08	.00	.28	.38*	.06	-.02	-.32	.16
24.	Emotional_Support (Village@School)_measurement_occasion_1	.21	.04	.21	-.24	-.11	-.12	-.17	-.00	-.11	.23	.22	-.25	-.01
25.	Classroom_Organisation (Village@School)_measurement_occasion_1	.10	-.16	-.01	-.29	-.03	.04	-.11	.12	.06	-.07	-.04	-.17	-.01
26.	Instsructional_Support (Village@School)_measurement_occasion_1	.09	-.12	.16	-.16	-.15	.05	.07	.43*	.33	-.05	.04	-.15	.20
27.	Positive Climate (Village@School)_measurement_ occasion2	.37	-.08	-.04	-.27	.34	.12	-.03	-.09	.12	.14	.13	.06	.05
28.	Teacher Sensitivity (Village@School)_measurement_occasion 2	.23	.17	.02	-.14	.05	.04	-.25	-.04	.16	.32	.23	-.05	.06
29.	Regard_For_Student_Perspectives (Village@School)_measurement_occasion 2	.20	-.19	-.03	-.06	.14	.08	.02	.08	-.15	.10	-.02	.06	.03
30.	Behavior_Management (Village@School)_measurement_occasion 2	-.23	-.17	-.21	.30	.03	.01	-.18	-.11	-.13	-.02	-.30	.19	-.13
31.	Productivity (Village@School)_measurement_occasion 2	-.20	-.01	.26	.21	.11	-.18	.28	-.01	-.03	-.03	-.01	.11	.07
32.	Negative_Climate_Rev (Village@School)_measurement_occasion 2	.32	-.01	-.08	-.12	.19	.14	.01	.09	.34	.15	.13	.08	.19
33.	Content_Understanding (Village@School)_measurement_occasion_2	.12	-.03	-.31	-.09	.25	.30	.22	.08	-.08	-.29	-.09	.19	.01
34.	Analysis_and_Inquiry (Village@School)_measurement_occasion_2	-.07	-.29	-.11	-.26	-.05	.10	-.10	-.06	-.03	-.06	-.26	-.14	-.14
		1	2	3	4	5	6	7	8	9	10	11	12	13

36. Instructional_Dialogue (Village@School)_measurement_occasion_2	.05	-.17	.05	-.03	.14	.10	-.10	.15	.16	.29	-.06	.09	.16
37. Emotional_Support (Village@School)_measurement_occasion_2	.35	-.03	-.02	-.22	.25	.11	-.11	-.04	.09	.24	.15	.03	.06
38. Classroom_Organisation (Village@School)_measurement_occasion_2	.09	-.08	-.05	.10	.19	.05	.04	.03	.20	.10	-.02	.18	.12
39. Instructional_Support (Village@School)_measurement_occasion_2	.07	-.17	-.14	-.18	.20	.15	-.05	-.00	.03	.09	-.12	.06	.01
40. Positive_Climate_post	.10	.25	-.24	.06	-.06	.15	.08	.23	.31	-.01	.09	.06	.26
41. Teacher_Sensitivity_post	.18	.45*	.03	.12	.46*	.47*	.39	.34	.26	.17	.37	.48*	.48*
42. Regard_for_Student_Perspectives_post	-.04	-.16	-.04	-.03	-.21	-.44*	-.52**	-.16	-.12	-.24	-.14	-.29	-.43*
43. Behavior_Management_post	-.24	.31	-.20	.06	.12	.38	-.04	.25	.31	.34	-.05	.24	.29
44. Productivity_post	.12	.22	.33	.31	.10	.01	.24	.20	-.01	.19	.33	.24	.24
45. Negative_Climate_Rev_post	.16	.03	-.13	.14	.09	.25	.36	.15	.36	-.03	.06	.23	.36
46. Content_Understanding_post	.33	.28	-.18	.00	.37	.45*	.28	.41*	.55**	-.13	.27	.36	.37
47. Analysis_and_Inquiry_post	.12	-.09	.16	-.05	.04	.33	.34	.37	.39	.02	.08	.12	.35
48. Quality_of_Feedback_post	.11	.09	-.15	-.13	-.21	.27	.10	.50*	.52**	-.09	.05	-.08	.32
49. Instructional_Dialogue_post	-.02	-.24	-.09	-.09	.01	.30	.17	.23	.56**	.02	-.19	.07	.31
50. Emotional_Support_post	.15	.38	-.17	.11	.14	.20	.07	.30	.32	-.00	.23	.21	.29
51. Classroom_Organisation_post	.02	.27	-.05	.23	.15	.35	.28	.29	.36	.23	.14	.35	.45*
52. Instructional_Support_post	.14	.10	-.10	-.03	.11	.47*	.27	.51*	.62**	-.01	.10	.22	.44*

*p < .05, **p < .01

Table 18. Bivariate correlations between Study Variables (CLASS-variables of measurement occasions Village@School/post with CLASS-variables occasion 1 Village@School)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Positive Climate (Village@School)_measurement_ occasion1	1												
2. Teacher Sensitivity (Village@School)_measurement_occasion 1	.21	1											
3. Regard_For_Student_Perspectives (Village@School)_measurement_occasion 1	.13	-.28	1										
4. Behavior_Management (Village@School)_measurement_occasion 1	-.07	.04	.11	1									
5. Productivity (Village@School)_measurement_occasion 1	.29	.22	.22	.10	1								
6. Negative_Climate_Rev (Village@School)_measurement_occasion 1	.28	-.08	-.25	.13	.13	1							
7. Content_Understanding (Village@School)_measurement_occasion_1	-.13	-.08	-.03	.32	.24	.08	1						
8. Analysis_and_Inquiry (Village@School)_measurement_occasion_1	-.23	.15	.14	.28	.27	-.06	.41*	1					
9. Quality_of_Feedback (Village@School)_measurement_occasion_1	.18	.11	.07	.34	.20	.13	.63**	.41*	1				
10. Instructional Dialogue (Village@School)_measurement_occasion_1	.09	-.03	.47**	.35	.44*	-.01	.44*	.58**	.56**	1			
11. Emotional_Support (Village@School)_measurement_occasion_1	.82**	.54**	.40*	.03	.41*	.02	-.14	-.00	.21	.27	1		
12. Classroom_Organisation (Village@School)_measurement_occasion_1	.28	.10	.04	.60**	.66**	.66**	.33	.25	.34	.40*	.25	1	
13. Instsructional_Support (Village@School)_measurement_occasion_1	-.03	.13	.17	.35	.35	-.00	.78**	.78**	.78**	.76**	.13	.36	1
14. Positive Climate (Village@School)_measurement_ occasion2	.39*	.24	.09	.17	-.05	.08	.16	-.04	.40*	.23	.42*	.10	.26
15. Teacher Sensitivity (Village@School)_measurement_occasion 2	.46*	.60**	-.02	.36	.18	-.01	-.08	-.01	.11	.13	.59**	.27	.09
16. Regard_For_Student_Perspectives (Village@School)_measurement_occasion 2	.43*	-.07	.37*	-.12	-.05	-.01	.03	-.26	.25	.22	.40*	-.08	.02
17. Behavior_Management (Village@School)_measurement_occasion 2	.21	-.18	-.07	-.13	-.18	.28	-.13	-.15	.23	-.12	.00	-.02	-.04
18. Productivity (Village@School)_measurement_occasion 2	.05	-.02	-.02	.06	.12	-.02	.21	.17	.09	.17	.01	.08	.26
19. Negative_Climate_Rev (Village@School)_measurement_occasion 2	.39*	.22	.17	.45*	.04	.01	.28	-.03	.45*	.37	.44*	.24	.33
20. Content_Understanding (Village@School)_measurement_occasion_2	.06	.06	-.28	.08	-.14	.28	.47**	.02	.36	-.14	-.06	.11	.25
21. Analysis_and_Inquiry (Village@School)_measurement_occasion_2	-.02	-.09	.13	.16	-.03	.43*	.37*	.22	.41*	.33	.01	.29	.41*
22. Quality_of_Feedback (Village@School)_measurement_occasion_2	.37*	.18	.12	.15	.14	.23	.23	-.06	.49**	.27	.38*	.27	.30
23. Instructional_Dialogue (Village@School)_measurement_occasion_2	.22	.03	.24	.27	-.09	.23	.15	-.09	.51**	.18	.26	.20	.20
24. Emotional_Support (Village@School)_measurement_occasion_2	.55**	.37*	.16	.21	.04	.04	.06	-.11	.35	.25	.61**	.15	.19
25. Classroom_Organisation (Village@School)_measurement_occasion_2	.39*	.09	.09	.32	.01	.10	.24	-.01	.46*	.30	.33	.21	.34
26. Instructional_Support (Village@School)_measurement_occasion_2	.21	.09	.06	.17	-.04	.33	.36	.02	.55**	.19	.21	.23	.36
27. Positive_Climate_post	.06	.50**	-.11	.22	-.19	-.06	-.01	-.01	.12	-.10	.26	-.04	.08
28. Teacher_Sensitivity_post	-.03	.15	.15	.38	-.23	-.14	.06	.00	-.03	-.03	.14	-.03	-.00
29. Regard_for_Student_Perspectives_post	.21	.20	-.32	-.09	.26	.28	.14	.17	.37	.10	.09	.27	.20
30. Behavior_Management_post	-.35	.15	-.12	-.00	-.20	-.15	-.33	-.15	-.23	-.21	-.20	-.20	-.25
31. Productivity_post	-.19	.05	.17	.03	-.02	-.26	.39	.22	.16	.01	-.01	-.14	.28
32. Negative_Climate_Rev_post	.18	.08	-.14	.10	-.23	.18	-.04	-.07	-.02	-.09	.09	.01	.00
33. Content_Understanding_post	-.39	-.09	-.15	.33	-.05	-.18	.62**	.29	.38	.20	-.37	.03	.42*
34. Analysis_and_Inquiry_post	-.18	.09	-.11	-.13	-.19	-.16	.41*	.28	.11	.14	-.12	-.27	.33

	1	2	3	4	5	6	7	8	9	10	11	12	13
35. Quality_of_Feedback_post	-.28	.07	.15	.11	-.05	-.27	.46*	.41*	.41*	.31	-.06	-.12	.49*
36. Instructional_Dialogue_post	-.34	-.11	-.11	-.06	-.22	.12	.26	.20	.15	.17	-.33	-.09	.23
37. Emotional_Support_post	.12	.54**	-.13	.34	-.16	-.00	.09	.07	.23	-.05	.31	.07	.14
38. Classroom_Organisation_post	-.15	.14	-.08	.07	-.24	-.07	-.05	-.04	-.07	-.16	-.06	-.15	-.02
39. Instructional_Support_post	-.41*	-.01	-.08	.17	-.17	-.17	.55**	.35	.32	.22	-.30	-.11	.44*

*p < .05, **p < .01

Table 19. Bivariate correlations between Study Variables (CLASS-variables of measurement occasion 2 Village@School/post with CLASS-variables occasion 2 Village@School)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Positive Climate (Village@School)_measurement_ occasion2	1												
2. Teacher Sensitivity (Village@School)_measurement_occasion 2	.45*	1											
3. Regard_For_Student_Perspectives (Village@School)_measurement_occasion 2	.44*	.01	1										
4. Behavior_Management (Village@School)_measurement_occasion 2	.31	.01	.27	1									
5. Productivity (Village@School)_measurement_occasion 2	-.03	-.03	-.10	.17	1								
6. Negative_Climate_Rev (Village@School)_measurement_occasion 2	.63**	.65**	.40*	.09	.08	1							
7. Content_Understanding (Village@School)_measurement_occasion_2	.26	.01	.01	.07	-.02	.05	1						
8. Analysis_and_Inquiry (Village@School)_measurement_occasion_2	.40*	.13	.16	.32	.12	.17	.51**	1					
9. Quality_of_Feedback (Village@School)_measurement_occasion_2	.79**	.33	.38*	.33	.03	.45*	.49**	.57**	1				
10. Instructional_Dialogue (Village@School)_measurement_occasion_2	.51**	.34	.49**	.37*	-.08	.63**	.22	.55**	.61**	1			
11. Emotional_Support (Village@School)_measurement_occasion_2	.90**	.69**	.59**	.27	-.06	.76**	.15	.33	.71**	.60**	1		
12. Classroom_Organisation (Village@School)_measurement_occasion_2	.57**	.50**	.33	.50**	.52**	.82**	.04	.31	.46*	.58**	.64**	1	
13. Instructional_Support (Village@School)_measurement_occasion_2	.64**	.29	.36	.39*	.06	.43*	.67**	.82**	.86*	.76**	.60**	.48**	1
14. Positive_Climate_post	.15	.45*	.21	-.03	-.22	.38	.12	-.07	.08	.25	.36	.19	.13
15. Teacher_Sensitivity_post	.39*	.29	.30	-.06	.01	.41*	.09	.15	.18	.26	.44*	.31	.20
16. Regard_for_Student_Perspectives_post	-.06	.02	-.27	.05	-.06	-.13	.25	.11	.04	-.09	-.10	-.11	.12
17. Behavior_Management_post	-.10	-.08	.25	.27	-.21	-.11	-.24	-.17	-.01	.11	-.02	-.08	-.09
18. Productivity_post	.01	-.01	.30	-.12	.34	.21	-.05	.05	-.07	.21	.08	.25	.13
19. Negative_Climate_Rev_post	.20	.23	.26	.18	-.08	.42*	-.05	-.10	.07	.24	.30	.37	.04
20. Content_Understanding_post	.13	-.01	-.07	-.13	-.08	.23	.39	.25	.05	.07	.05	.10	.19
21. Analysis_and_Inquiry_post	-.03	-.24	.16	-.28	-.01	-.04	.37	.11	-.09	-.06	-.09	-.15	.06
22. Quality_of_Feedback_post	-.06	-.12	.16	-.25	-.38	.16	.14	.01	-.13	.11	-.05	-.13	.01
23. Instructional_Dialogue_post	.23	-.23	.30	.26	-.13	.08	.15	.36	.18	.24	.11	.10	.26
24. Emotional_Support_post	.32	.51**	.21	-.04	-.17	.45*	.24	.09	.18	.29	.48*	.27	.27
25. Classroom_Organisation_post	.07	.09	.40*	.21	-.03	.26	-.18	-.13	.01	.28	.19	.27	.03
26. Instructional_Support_post	.05	-.14	.15	-.13	-.17	.18	.29	.21	-.04	.15	.00	.02	.14

*p < .05, **p < .01

Table 20. Bivariate correlations between Study Variables (CLASS-variables of post-measurement)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Positive_Climate_post	1												
2. Teacher_Sensitivity_post	.42*	1											
3. Regard_for_Student_Perspectives_post	-.18	-.57**	1										
4. Behavior_Management_post	.45*	.17	-.34	1									
5. Productivity_post	.18	.43*	-.34	-.06	1								
6. Negative_Climate_Rev_post	.72**	.35	-.38*	.38*	.08	1							
7. Content_Understanding_post	.09	.32	.13	-.04	.21	.02	1						
8. Analysis_and_Inquiry_post	.27	.22	.05	.04	.18	.21	.47*	1					
9. Quality_of_Feedback_post	.47*	.12	.13	.19	.16	.29	.58**	.61**	1				
10. Instructional_Dialogue_post	.13	.16	-.02	.26	-.03	.37	.50**	.55**	.50**	1			
11. Emotional_Support_post	.88**	.65**	-.03	.27	.24	.55**	.31	.34	.46*	.18	1		
12. Classroom_Organisation_post	.72**	.46*	-.53**	.72**	.39*	.81**	.07	.21	.32	.34	.55**	1	
13. Instructional_Support_post	.34	.34	.03	.21	.22	.30	.81**	.77*	.84**	.75**	.46*	.37	1

*p < .05, **p < .01

When looking in more detail to the teacher style over the four measurement occasions (see Table 21, 22, 23 and 24), results indicate that the same three CLASS dimensions show a significant evolution over the course of the trajectory. In Table 9 the variances at the school, class and time level are given, for the baseline models as well as for the models with time as predictor. While at first the variance at the class level for the Quality of Feedback dimension was significant when just adding the predictor time, significant differences between classes in the mean Quality of Feedback were no longer found after we allowed the predictor time to vary across classes.[53] However, classes seem to differ in their growth in Quality of Feedback. When the predictor time was added to the models of Regard for Student Perspectives, Quality of Feedback and Negative Climate (reversed), the deviance of the total models reduced significantly (X^2 (1) = 9.15** for Regard for Student Perspectives, X^2 (1) = 8.72 for Negative Climate (reversed) and X^2(1) = 5.07* for Quality of Feedback).

When looking to the intercepts after adding the different centered time predictors (measurement occasions 1, 2, 3 and 4) to the models with the dimensions in which significant growth was detected, it becomes clear where precisely the growth is situated. For Quality of Feedback, there is a growth in all measurement occasions (4,15->4,27->4,40->4,53); for Negative Climate (Reversed), there is a decrease over all measurement occasions (6.80->6.64->6.48->6.31) and for Regard for Student Perspectives, there is an increase over all measurement occasions (4.6->4.74->4.88->5.01). This means that also during Village@School teachers tended to grow in these dimensions.

[53] This may indicate that these models are not stable. The model in which we allowed the predictor time to vary across classes is not shown in Table 9 because it had the lowest badness of fit in comparison to the baseline model. Only the model with the highest badness of fit (in which we didn't allow the time predictor to vary across classes) is shown in Table 9.

Table 21: Predicting the CLASS domains by the Predictor Time

	Emotional Support (n=116)[a]								**Classroom Organisation (n=115)**						**Instructional Support (n=116)**								
	Baseline Model[b]		Model A1 (time not allowed to vary across classes)			Model A2 (time allowed to vary across classes)			Baseline Model			Model A3 (time not allowed to vary across classrooms)			Baseline Model			Model A4 (time not allowed to vary across classrooms)			Model A5 (time allowed to vary across classrooms)		
	B	SE	B	SE	Δ R2	B	SE		B	SE		B	SE	Δ R2	B	SE		B	SE	Δ R2	B	SE	
Random parameters					-.01									.00						-.01			
Class level																							
σ_{cons}^2	.04*	.02	.05*	.02		-.00	.03								.06*	.03		.07*	.03		-.01	.05	
$\sigma_{cons.time}$						.02*	.01														.04	.02	
σ_{time}^2						-.01	.01														-.02	.01	
Time level																							
σ_{cons}^2	.11***	.02	.11***	.02		.12***	.02		.14***	.02		.14***	.02		.20***	.03		.19***	.03		.23	.04	
Deviance			102.331			98.805			100.256			98.232			167.776			164.691			161.309		
					β			β			β			β			β			β			β
Fixed parameters																							
Intercept			4.90***	.06		4.90***	.05		5.89***	.04		5.96***	.06		4.58***	.06		4.48*	.08		4.49***	.07	
Time **1.12**			.04	.03	.11	.04	.02	.11				-.05	.03	-.15				.07	.04	.15	.07*	.03	.15*

[a] In the prediction of Emotional Support, Classroom Organisation and Instructional Support, data were available for 34 teachers, yielding (34 x 4 occasions =) 136. Over the 4 occasions, 20 pieces of information were missing for Emotional Support and Instructional Support, and 21 pieces of information were missing for Classroom Organisation.

[b]The baseline model did not include any predictors.

*p < 0.05, **p < 0.01; *** p < 0.001

Table 22: Predicting the dimensions of the Emotional Support domain by the Predictor Time

	Positive Climate (n=116)[a]								Teacher Sensitivity (n=115)								Regard for Student Perspectives (n=116)					
	Baseline Model[b]		Model A6 (time not allowed to vary across classes)			Model A7 (time allowed to vary across classes)			Baseline Model		Model A8 (time not allowed to vary across classrooms)			Model A9 (time allowed to vary across classes)			Baseline Model			Model A10 (time not allowed to vary across classrooms)		
	B	SE	B	SE	Δ R2	B	SE		B	SE	B	SE	Δ R2	B	SE		B	SE		B	SE	Δ R2
Random parameters					.00								.00									.03
Class level																						
σ_{cons}^2	.11*	.06	.11*	.06		.04	.10		.12*	.05	.12*	.05		.08	.08							
$\sigma_{cons.time}$						.04	.04							.02	.03							
σ_{time}^2						-.02	.02							-.01	.01							
Time level																						
σ_{cons}^2	.37***	.06	.37***	.06		.40***	.08		.23***	.04	.23***	.04		.25***	.05		.30***	.04		.27***	.04	
Deviance	236.963		236.915			235.953			192.545		192.540			191.709			188.144			178.997		
Fixed parameters					β			β					β			β			β			β
Intercept	4.91***	.08	4.92***	.11		4.92***	.10		5.14***	.08	5.15***	.10		5.15***	.09		4.81***	.05		4.61***	.08	
Time			-.01	.05	-.02	-.01	.05	-.02			-.00	.04	.00	-.00	.04	.00				.14**	.04	.29**

[a] In the prediction of Positive Climate, Teacher Sensitivity and Regard for Student Perspectives, data were available for 34 teachers, yielding (34 x 4 occasions =) 136. Over the 4 occasions, 20 pieces of information were missing for Positive Climate, Teacher Sensitivity and Regard for Student Perspectives.

[b]The baseline model did not include any predictors.

*$p < 0.05$, **$p < 0.01$; *** $p < 0.001$

Table 23: Predicting the dimensions of the Classroom Organisation domain by the Predictor Time

	Behavior Management (n=116)[a]					**Negative Climate Rev. (n=115)**					**Productivity (n=116)**				
	Baseline Model[b]		Model A11 (time not allowed to vary across classes)			Baseline Model		Model A12 (time not allowed to vary across classrooms)			Baseline Model		Model A13 (time not allowed to vary across classrooms)		
	B	SE	B	SE	Δ R2	B	SE	B	SE	Δ R2	B	SE	B	SE	Δ R2
Random parameters					.00					.03					.00
Class level															
σ_{cons}^2															
$\sigma_{cons.time}$															
σ_{time}^2															
Time level															
σ_{cons}^2	.30***	.04	.30***	.04		.44***	.06	.41***	.05		.26***	.03	.26***	.03	
Deviance	190.105		190.092			231.884		223.167			173.047		172.266		
Fixed parameters					β					β					β
Intercept	5.67***	.05	5.68***	.09		6.56***	.06	6.80***	.10		5.45***	.05	5.39***	.08	
Time			-.01	.05	-.02			-.16**	.05	-.27**			.04	.04	.09

[a] In the prediction of Behavior Management, Negative Climate (Rev.) and Productivity, data were available for 34 teachers, yielding (34 x 4 occasions =) 136. Over the 4 occasions, 20 pieces of information were missing for Behaviour Management and Productivity, and 21 pieces of information were missing for Negative Climate (Reversed).

[b]The baseline model did not include any predictors.

*p < 0.05, **p < 0.01; *** p < 0.001

Table 24: Predicting the dimensions of the Instructional Support domain by the Predictor Time

	Content Understanding (n=116)[a]								**Analysis and Inquiry (n=116)**					**Quality of Feedback (n =116)**								**Instructional Dialogue (n=116)**				
	Baseline Model		Model A14 (time not allowed to vary across classes)			Model A15 (time allowed to vary across classes)			Baseline Model		Model A16 (time not allowed to vary across classes)			Baseline Model		Model A17 (time not allowed to vary across classes)			Model A18 (time allowed to vary across classes)			Baseline Model		Model A19 (time not allowed to vary across classes)		
	B	SE	B	SE	Δ R2	B	SE		B	SE	B	SE	Δ R2	B	SE	B	SE	Δ R2	B	SE		B	SE	B	SE	Δ R2
Random parameters					.01								.01					.01								.00
Class level																										
σ_{cons}^2	.20*	.09	.21*	.09		.16	.15							.14*	.07	.15*	.07		-.11	.10						
$\sigma_{cons.time}$						.03	.05												.13***	.04						
σ_{time}^2						-.02	.03												-.07**	.02						
Time level																										
σ_{cons}^2	.49***	.08	.47***	.07		.50***	.09		.61***	.08	.60***	.08		.45***	.07	.42***	.06		.53	.10		.41***	.05	.41***	.05	
Deviance	274.143		270.890			270.384			271.013		270.215			258.848		254.781			244.249			226.226		226.072		
					β			β					β					β			β					β
Fixed parameters																										
Intercept	4.06***	.10	3.91***	.13		3.91***	.13		4.50***	.07	4.41***	.12		4.33***	.09	4.15***	.12		4.15***	.09		4.75***	.06	4.71	.10	
Time			.11	.09	.15	.11*	.05	.15*			.06	.07	.09			.13*	.06	.19*	.13**	.04	.19*			.02	.05	.04

[a] In the prediction of Content Understanding, Analysis and Inquiry, Quality of Feedback and Instructional Dialogue, data were available for 34 teachers, yielding (34 x 4 occasions =) 136. Over the 4 occasions, 20 pieces of information were missing for these Instructional Support dimensions.

[b]The baseline model did not include any predictors.

*p < 0.05, **p < 0.01; *** p < 0.001

Conclusions and discussion

It is a challenge for teachers to implement open-ended S&T learning environments. First, we explored whether the teacher's competence profile, in particular his/her attitudes towards S&T (teaching) and his/her teacher style in such learning environments, plays a role for his/her teacher style while implementing the challenging project-based learning environment Village@School. Teacher style was operationalised as the quality of the teacher-pupil and pupil-pupil interactions. From the literature it is well known that negative attitudes towards science exist among primary school teachers (Gustafson & Rowell, 1995; McDuffie, 2001; Parker & Spink, 1997; Palmer, 2001; Skamp & Mueller, 2001). Moreover, teachers' attitudes, knowledge and skills have an influence on the way in which new curricula are implemented (Krajcik et al., 1994; Rogers et al., 2011). Second, we investigated whether teachers grew in their competence profile because of the implementation of Village@School and the support that was provided. For both teachers' attitudes towards S&T (teaching) and their teacher style, the growth after the project in comparison to before was determined. To get insight in the process of the evolution of teacher style, it was also investigated whether teachers grew during the implementation of the project itself.

Most importantly with regard to research question 1, we first found that the differences between schools in the level of emotional support that was provided during Village@School can be explained by the emotional support provided by teachers in open-ended S&T activities before the project. This means that schools having a more positive climate in their classes during the S&T assignment before the project, and which have teachers, who show more sensitivity, responsiveness and openness for students' ideas during this assignment, have a higher emotional support during the Village@School project (when controlling for the other attitudes and CLASS-domains). However, and second, teachers' attitudes towards inquiry learning and their initial classroom organisation, are negatively related to the emotional support that was provided during Village@School when controlling for the other attitudes and CLASS domains. One explanation could be that teachers who have a more positive attitude towards inquiry learning – which is characterised by an openness for pupils' experimentation – may have too high expectations for pupils, which may result in teachers being less emotionally supportive during Village@School. When it comes to high-quality S&T education, the literature points to learning environments in which students are given high levels of autonomy (e.g. Barak & Doppelt, 2000; Furtak et al., 2012). Though the project aims at stimulating collaboration among pupils, teachers still need to provide support. In another study (see Study 2), we showed that teachers' sensitivity and responsiveness for pupils' needs is important for the latter's growth in engagement in the investigated open-ended S&T learning environments. Furthermore, teachers who have an initial higher classroom organisation – which is part of their teacher style and characterised by a smooth organisation of class processes, a good management of student behaviour and less negativity - may experience more stress to guarantee productivity and manage pupils' behaviours in the more challenging and longer taking project Village@School. Inquiry- and design-based learning environments are more complex and difficult to manage forms of instruction (Keys & Bryan, 2001), and this may especially come to the surface in Village@School. It is plausible that the higher stress with regard to this classroom organisation of teachers initially scoring higher on classroom organisation, results in a diminished emotional support while implementing Village@School.

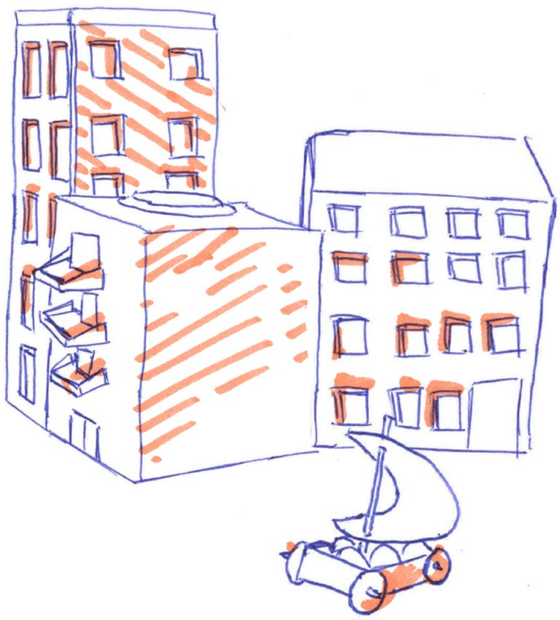
Still, our findings are surprising, as we expected the four teacher attitudes to be positively related to classroom interactions during the Village@School project. This finding can be explained in several ways. First, the literature indicates that powerful inquiry learning environments may require more than teachers' initial attitudes towards S&T (teaching) and classroom interactions in inquiry- and design-based learning environments alone. The implementation of inquiry lessons by science teachers is influenced by a multitude of factors (Roehrig & Luft, 2004). Also a teacher's content knowledge (Gess-Newsome, 1999) and pedagogical content knowledge (Crawford, 2000) are important determinants of high quality inquiry based teaching. Second, Mansour (2013) described that teacher practices only partially corresponded to their beliefs and attitudes. Third, as our sample teachers consciously chose to participate in the challenging project Village@School, they may vary only slightly, both with regard to their attitudes towards S&T (teaching) and their teacher style. Only for emotional support during Village@School we found a significant variance at school level. The latter explains why more relations between teachers' competence profile and CLASS domains could not be detected.

Concerning the second research question with regard to teachers' competence profile, we did not find a significant positive change in teachers' attitudes towards S&T (teaching) as a result of the implementation of Village@School and the support provided during the project. Some other researchers did also not report significant changes in attitude over time as a result of practical experience with inquiry-based science teaching (e.g. Kang, 2007). Different explanations can be given. First, as argued before, implementing the Village@School project is challenging for teachers, as they have to give much initiative to students, while still maintaining an active role. The teachers were supported throughout the project – via a start conference, two workshops with the whole sample group, and two individual coaching sessions – but because of practical reasons, i.e. intensive data collection in a quite large group of teachers, this support may have not been effective for some teachers, causing no change in their attitudes. The amount of time spent on this training was smaller than the 20 to 50 contact hours that are recommended by several scientific guidelines for effective professional development (Blank, de las Alas, & Smith, 2008; Borko, 2004). Furthermore, the training that was provided mainly focused on the implementation of Village@School, rather than on the teacher attitudes with regard to S&T (teaching). Van Aalderen-Smeets et al. (2015) found that in order to stimulate primary school teachers' attitudes towards science it is important to raise awareness about their own attitudes towards teaching science, to let them reflect on these attitudes, and to show them how to teach science independently of specific science content or pre-structured teaching materials. In our study, teachers' self-reflection on what they were doing was limited, which may be a reason a growth in attitude was not reflected. Second, because the sample consisted of a group of teachers who voluntarily chose to participate in this study, their attitudes were already rather high at the start (for each of the four attitudes the mean was above the mean of the 5-point Likert scale). As a consequence of this ceiling effect, opportunities to grow were limited.

Also with regard to our second research question, we found a significant positive change in Regard for Student Perspectives and Quality of feedback. Even though we were not able to use a control group, there are good reasons to claim that the growth in these dimensions may be attributable to the implementation of the innovative learning environment and to the training that we provided to teachers. Teachers were at first instance trained in the basic principle of the project, i.e. showing an openness for pupils' ideas and giving them responsibility while constructing the miniature site. In combination with this, the importance of open, stimulating questions that make pupils think was stressed on a regular basis. However, our contacts with the participating teachers taught that teachers struggled with finding a balance between giving not enough or too much initiative. Some teachers were irritated when pupils didn't follow the way of planning, experimenting or designing that they had in mind. Teachers can no longer control what is learnt and how it is learnt. In Village@School the curriculum 'emerges' and goals cannot be determined beforehand. The project also requires patience from teachers as it can take a while before real, good constructions become visible on the standard plate. Moreover, some pupils struggled with finding the necessary materials and information to conduct their experiments and make their constructions. Teachers regularly felt powerless to help pupils because they experienced not having the necessary stock-in-trade. This may explain our second finding, namely the increase in Negative Climate, which manifests in irritation and anger of the teachers/pupils, over the course of the trajectory. We remark that the negative climate as evaluated after the project was still situated at the lower end of the Negative Climate scale of the CLASS.

Although this study is unique in that an intensive data collection via observations was conducted for a quite large group of teachers, a first limitation is that the number of participants is still rather small, which lowers the statistical power of our analyses. This may not only have had its consequences for the retrieved factors in the factor analyses of the attitude questionnaire, but may have also resulted in the limited effects of teachers' competence profile on teacher style during the implementation of Village@School. Second, the use of a control group in this study would have provided more evidence for the effect of the project on teacher style in S&T activities. Third, the literature shows that the teacher's competence profile consists of more elements than attitudes and skills alone (du Chatenier, Verstegen, Biemans, Mulder, & Otma, 2010; Mulder, 2001). Also teachers' knowledge, and particularly their subject matter knowledge (Abell, 2007; Gess-Newsome, 1999, in Friedrichsen, Van Driel, & Abell, 2010) and pedagogical content knowledge (Abell, 2007; Davis, Petish, & Smithy, 2006), may have played a role in the relations that were investigated in our two research questions. Finally, it has to be noticed that studying the change in teacher style between the pre- and post-measurement has to be treated as more 'pure', in comparison to investigating the evolution of the teacher throughout the whole intervention (pre, in-between, post), as for the pre- and post-measurement two standardised S&T-assignments could be created. Because of their short duration and the specific instructions (e.g. materials) provided, these assignments differ from the innovative curriculum of Village@School. However, Village@School is – just like the S&T-assignments – open-ended in nature. The advantage of our investigation of the growth in-between the pre- and post-measurement is that it opens the black box of the intervention. The investigation of this black box, together with the exploration of the relation between teacher style before the project and teacher style during Village@School, taught us more about the relative stability of the teacher style. Teacher style in the innovative and challenging project-based learning environment was only partly, and not necessarily positively, determined by the teacher style in another open-ended S&T activity, and teacher style could evolve throughout the project with regard to some aspects. Results with regard to the latter showed to be the same as the results with regard to the change in teacher style, based on the difference in teacher style in the two standardised S&T assignments.

Future research can build further on the findings and (hypothesised) explaining factors of this study. First, it is recommended to profoundly investigate whether teachers' insecurity about teaching S&T in learning environments in which pupils are permitted a large amount of initiative, plays a role in the quality of the interactions that are an indication of the teacher's style. This emotional support may be especially important in this type of learning environments, as pupils are not used to working autonomously and need a sensitive teacher who shows recognition and supports their work in order for their engagement to grow. Another message of this study is that teachers may improve with regard to the autonomy they give to pupils and improve the atmosphere of giving feedback that aims for a deep level of understanding in pupils. By that we met the call of Van Aalderen-Smeets & Walma van der Molen (2015) for professional development to focus on the improvement of the competency of teachers to adjust their teaching to the students instead of on the teaching of a certain 'proven' method or content. A recent review study on the effects of primary science teaching methods on student outcomes shows that providing teachers with science kits does not affect student outcomes (Slavin, Lake, Hanley, & Thurston, 2014). However, another important conclusion of Van Aalderen-Smeets and Walma van der Molen (2015) was that the most effective way to obtain attitude changes in teachers is to create awareness and reflection, not only upon one's attitudes and beliefs, but also upon one's emotions and upon one's teaching and behaviour. Furthermore, an increase in the attitude of teachers was found after an intervention similar to ours in which a visit to a science centre was also included (Van Cleynenbreugel, De Winter, Buyse, & Laevers, 2011). The visit was important for teachers' attitude development as lot of inspirational materials were provided for teachers. Teachers also learnt that a rich environment is an important basis for pupils to learn about S&T in co-construction. In future research, it will be important to create more opportunities to guarantee this reflection.



Study 4: Comparing the Adult Style Observation Schedule (ASOS-E) and the Classroom Assessment Scoring System Upper Elementary (CLASS): A Concurrent Validity Study

Submitted

Keywords: teacher style, teacher-student interactions, Adult Style Observation Schedule (ASOS), Classroom Assessment Scoring Sytem (CLASS)

Abstract

In the present study, the latest version of the Adult Style Observation Schedule (ASOS-E) was validated, by means of convergent validation techniques. The associations between the dimensions of the ASOS were examined in relation to relevant dimensions of the Classroom Assessment Scoring System Upper Elementary (CLASS). Both measures were applied to 22 teachers of the 4th, 5th and 6th grade of primary school while they conducted inquiry- and design-based activities in the fields of science and technology (over a total of 30 observations). Correlations show that convergent validity of ASOS and CLASS exists for some of the dimensions. In particular, congruence was found between the ASOS dimension Stimulation and the Content Understanding and Positive Climate dimensions of the CLASS. For the ASOS dimension Sensitivity, a connection with the CLASS dimension Positive Climate was found. No evidence emerged for the convergence of the Giving Autonomy dimension of the ASOS with similar CLASS dimensions. Implications for future research are discussed.

Introduction

The way in which the teacher interacts with learners – called by some the teacher's style – characterises the teacher's generalised response tendency (Pettigrew et al., 2013). The patterned ways in which the teacher interacts with students (Pettigrew et al., 2013) has been shown to be important for students to be involved and to learn (Edmunds et al., 2008; Walker, 2008). In this study we examined an existing instrument that measures the concept of teacher style. More specifically, the validity of the Adult Style Observation Schedule (ASOS) (Laevers & Heylen, 2013), as developed at the Centre for Experiential Education (CEGO), was explored. Before discussing the instrument and its validity, a brief overview of the literature with regard to teacher style is provided.

Theoretical Framework

Teacher Style

The concept of teacher style is mainly described as the general stance of the teacher towards his or her students (Laevers, 2005; Sierens et al., 2006; Sweertvaegher, 2008) or as the individual pattern of the teacher in the way in which he or she intervenes in a wide variety of situations (Laevers, 2005; Pettigrew et al., 2013). Some, but not all, conceptualisations also emphasise that teacher style becomes apparent in the interactions between teacher and learners (de Kruif et al., 2000; Laevers & Heylen, 2013). Other authors point to the importance of interactions between teacher and learners but do not use the concept of teacher style (e.g. Pianta et al., 2012). Finally, still other educational researchers use the concept of 'interpersonal style' (Deci, Schwartz, Sheinmann, & Ryan, 1981; Vallerand, Fortier, & Guay, 1997).

The most prominent theory pointing to the importance of teaching style – or interpersonal style – is self-determination theory (SDT; Deci & Ryan, 2000). In this theory, learners' motivation and self-regulated learning are facilitated by nurturing their basic psychological needs. These are the need for autonomy (i.e. experiencing a sense of volition), the need for competence (i.e. experience efficacy), and the need for relatedness (i.e. feeling connected to others). According to SDT, students' perceptions of their teachers as need-supportive or need-frustrating have important consequences for their motivation and subsequent learning strategies and achievement (Niemiec & Ryan, 2009; Vansteenkiste, Zhou, et al., 2005). More specifically, teaching that supports students' needs fosters more autonomous motivation, in comparison to teaching that does not meet these needs (Niemic & Ryan, 2009). Autonomous motivation refers to learning in a volitional way because of the perceived value or inherent satisfaction of the learning activity (Ryan & Connell, 1989). This motivation is different from a controlled motivation, which deals with learning to meet external (e.g. rewards, external regulation) or internal (e.g. feelings of guilt, introjected regulation) pressures (Ryan, Connell, & Grolnick, 1992, in Soenens et al., 2012). It is repeatedly shown that an autonomous (as compared to a controlled) motivation is positively related to students' use of effective and thorough strategies of self-regulated learning, such as time management and deep-level cognitive processing (Vansteenkiste et al., 2009).

In the literature, teacher style is generally categorised using similar dimensions. Firstly, in the SDT, Deci and Ryan (2000) distinguish three dimensions of teaching style: Autonomy Support, Structure and Engagement. The starting point for these dimensions is the needs of the learners: their need for autonomy, for feeling competent and for feeling connected to significant others (Sierens et al., 2009). Autonomy Support deals with teachers' promotion of volitional functioning and fostering of a sense of initiative and interest in students (Soenens et al., 2012). The Structure dimension comprises giving information such that tasks can be accomplished and learning goals can be reached (Sierens et al., 2006). Finally, the Engagement dimension refers to the responsiveness of teachers, or their ability to make students feel relationally connected (Skinner & Belmont, 1993). Some researchers have conceptualised the styles of teachers in line with parenting styles, and use the concepts of 'authoritarian', 'permissive' and 'authoritative' teachers (e.g. Paulson et al., 1998; Wentzel, 2002), as proposed in Baumrind's (1973, in Pettigrew et al., 2013) taxonomy. In this view, a combination of the amount of control and responsiveness which are provided by the teacher leads to a characterisation of their style. Authoritarian teachers have moderate to high levels of control but low levels of responsiveness; permissive teachers show low control and low responsiveness; and authoritative teachers have moderate to high levels of control and high levels of responsiveness.

The Adult Style Observation Schedule (ASOS)

Researchers at the Centre for Experiential Education (CEGO) have developed an instrument to measure the concept of teacher style: the Adult Style Observation Schedule (ASOS) (Laevers & Heylen, 2013) (see Figure 1). In experiential education theory, teacher style – along with the provided materials and activities – is viewed important to achieve high levels of well-being and involvement in learners (Laevers et al., 2005). Teacher style can be observed through the interventions of the teacher and in the way the teacher deals with the learners (Laevers & Heylen, 2003; Laevers & Heylen, 2013). Their pattern of interacting is in line with their capability to take perspective (Benoit, 2015; Laevers & Heylen, 2003), which is the conscious focus of the teacher on what is happening in learners and in themselves (Laevers, Heylen, & Daniëls, 2004). The concept of 'modus' refers to the characteristics of the style, like 'lively', 'challenging', 'motivating' and so on, that contribute to learners' well-being and engagement (Benoit, 2015; Lento, 2016). The most recent version of the instrument, the ASOS-E, serves as an observation tool to understand how teachers intervene during educational activities taking place in care settings to higher education (Laevers & Heylen, 2013).

FIGURE 1. The Adult Style Observation Schedule (ASOS- E) with its dimensions, heads and indicators (Benoit, 2015)

STIMULATION

		INDICATORS (negative)	INDICATORS (positive)
HEADS	**INTRODUCING ACTIVITIES**	Introduces activities in a non-motivating way	Introduces activities in a motivating way
	PROVIDING INFORMATION	Tells or presents information in a way that is not appealing	Tells or presents information in a captivating way
	STIMULATING COMMUNICATION	Gives no or non-effective impulses to communicate	Invites to communicate in a stimulating manner
	STIMULATIING THINKING	Gives no thought-provoking impulses	Provokes thought by using stimulating questions and comments
	STIMULATING ACTION	Gives no or non-effective impulses to come into action	Gives challenging impulses to come into action
		Does not intervene when children are not active	Gives suggestions for activities where needed

SENSITIVITY

		INDICATORS (negative)	INDICATORS (positive)
HEADS	**RESPECT**	Puts her or himself above the children; does not show respect	Takes the children seriously; shows respect
	ATTENTION	Ignores the need for attention	Satisfies the need for attention
	AFFECTION	Shows no acceptance; has a cold and distant attitude	Shows acceptance, warmth and/or affection
	AFFIRMATION	Does not affirm capabilities; discourages	Affirms capabilities and empowers
	CLARITY	Does not care to clarify; provokes insecurity	Clarifies situations and makes children feel secure
	UNDERSTANDING	Shows no understanding of feelings or emotions	Shows sympathy for feelings and emotions

GIVING AUTONOMY

		INDICATORS (negative)	INDICATORS (positive)
HEADS	**INITIATIVE**	Rejects children's initiative and proposals; takes away opportunities for children to take initiative	Welcomes children's initiative and proposals; exploits opportunities for children to take initiative
	METHOD OF WORKING	Does not allow deviation from the planned procedure	Gives room to experiment, and to proceed in one's own way
	(FINAL) PRODUCT	Does not allow children to negotiate on the final result	Gives children responsibility for the final result
	RULES AND AGREEMENTS	Imposes rules and agreements	Involves children in the setting of rules and agreements
	CONFLICT	Does not involve children in solving conflicts	Involves children in the solving of conflicts

Figure 2. The minimum and maximum scores of the 3 ASOS-dimensions (Laevers, Berghmans, Declercq, & Buyse, October 2015)

		Very Low (1)	Very High (7)
DIMENSIONS	Stimulation	Acts most of the time in a habitual manner; does not capture children's interests; does not connect with their developmental level; takes away opportunities for meaningful communication, action and thinking; through all this, leaves an outspoken schoolish impression.	Intervenes with enthusiasm; has a particular sense for what speaks to children; is perfectly tuned into their level of development; is constantly provoking communication, action and thought and through all this, leaves a highly stimulating impression.
	Sensitivity	The interventions are dominated by a negative tone, a rejecting attitude and irritation, and an outspoken lack of empathy for the feelings and needs of children who are belittled, ignored, discouraged, feel insecure or not understood at all.	The interventions are highly positive and reflect an outspoken acceptance and empathy for the feelings and needs of children who are being respected, get attention, affection and affirmation, and feel secure and fully understood.
	Giving Autonomy	The autonomy of children is highly restricted. Expressed interests are being ignored, the standard procedures are strictly imposed, the final result must match the model, rules are being forced upon them, or conflicts are solved without any dialogue.	Children get ample room to manifest themselves. Expressed interests are welcomed, there is a lot of flexibility with regard to the procedures and final result, rules are openly discussed and conflicts resolved in dialogue.

The ASOS distinguishes three dimensions in teacher style – Stimulation, Sensitivity and Giving Autonomy – which are largely similar to the dimensions discerned in SDT. First, Stimulation interventions deal with the ways in which the teacher tries to raise learners' engagement during activities (Laevers & Heylen, 2013; Vervoort, 2011), such as "suggesting activities to children who wander around, offering materials that fit in an ongoing activity, inviting students to communicate, raising thought-provoking questions and giving information that can capture their mind" (Laevers, 2005, p.7). Not only the way in which activities are introduced and information is provided, but also whether and how communication, thinking and action are stimulated are part of this dimension. A teacher scoring high on the Stimulation dimension intervenes with enthusiasm, has a particular sense for what speaks to children, is perfectly tuned into their levels of development, is constantly provoking communication, action and thought and by doing all of this, leaves a highly stimulating impression (see Figure 2). Secondly, Sensitivity deals with showing respect, acceptance, warmth/affection and being sensitive and responsive to students' (emotional) needs (Benoit, 2015; Laevers & Heylen, 2013). The teacher shows understanding and provokes security for children (Benoit, 2015). Highly sensitive teacher's interventions are highly positive and reflect an outspoken acceptance and empathy for the feelings and needs of children: they are being respected, they get attention, affection and affirmation, and they feel secure and fully understood (see Figure 2). Thirdly, Giving Autonomy comprises the openness of the teacher towards students' initiatives and ideas, their own methods of working, learners determining the final result and students' participation when making rules/agreements and solving conflicts (Benoit, 2015; Laevers & Heylen, 2013). A teacher who scores highly for Giving Autonomy gives children ample room to manifest themselves (see Figure 2). Expressed interests are welcomed, there is a lot of flexibility with regard to the procedures and final result, rules are openly discussed and conflicts are resolved through dialogue.

The ASOS-E is an observation system which uses an open critical incident procedure (Benoit, 2015). The starting point of the observations are 'critical incidents', or events in the interactions between the teacher and learners which are relevant in light of one or more of the indicators of the different dimensions. More specifically, observed critical incidents are placed under different heads or aspects of which the three dimensions consist. These heads are then, in turn, further subdivided into indicators. The 'Introducing Activities' head for instance consists of the indicator "Introduces activities in a motivating/non-motivating way" (Benoit, 2015, p. 21). To determine whether interactions are 'critical' in light of a particular dimension, head or indicator, it is important to have a profound insight into these different elements of the observation scale.

Each head consists of a continuum with two extremes and two in-between scores. As such, a four-point scale (1 = not at all like, to 4 = very much like) is used to evaluate critical incidents on one or more of these heads. The observer writes the observed critical incident down under the right dimension and the fitting head and indicators (Lento, in preparation). Next, he/she assigns a score on the four-point scale to give a weight to the critical incident by ticking one out of four boxes (see Figure 3). When the observation cycle (limited to 10 or 15 minutes) is over, the observer gives a global episode score for each dimension on a 7-point scale. The minimum and maximum scores of these scales can be found in Figure 2. The global score is not the mathematical mean of the scores assigned to the registered indicators on the four-point scale, but a holistic score based on an evaluation in which the observer takes the pattern of teacher interventions with negative and positive loadings into account. The amount of critical incidents and the extent to which they are positively or negatively colored (as indicated on the four-point scale) are taken into consideration when assigning the global score for each dimension for the particular episode of a teacher. The observation system is 'open' because the observer is not forced to predetermined categories that are required to get a score. No score is given for the heads. The final judgement on the 7-point scale is based on the pattern of boxes that have been ticked and for which the indicators delivered the criteria.

FIGURE 3. Observation Sheet for the Stimulating dimension [Registration of critical incidents] (Laevers et al., October 2015)

WEAK – LOW	**STIMULATING**	**STRONG - HIGH**
•Introduces activities in a non-motivating manner	☐☐.................................☐☐ ☐☐.................................☐☐	•Introduces activities in a motivating manner
•Tells/brings information in a way that doesn't speak	☐☐.................................☐☐ ☐☐.................................☐☐	•Tells/brings information in a captivating way
•Gives no or non-effective impulses to communicate	☐☐.................................☐☐ ☐☐.................................☐☐	•Invites to communicate in a stimulating manner
•Gives no or non-effective impulses for action	☐☐.................................☐☐ ☐☐.................................☐☐	•Gives challenging impulses for action
•Gives no thought provoking impulses	☐☐.................................☐☐ ☐☐.................................☐☐	•Provokes thought by striking questions and comments
•Doesn't intervene when children are not active	☐☐.................................☐☐ ☐☐.................................☐☐	•Gives suggestions for activities where needed

The ASOS-E has certain features that distinguish it from other instruments. Firstly, while other instruments which measure teacher style are student questionnaires (Fraser, 1998; Reeve & Jang, 2006), the ASOS-E is an observation instrument. Secondly, the instrument is holistic, as critical incidents and the weight of these incidents in the foreseen observation period are used for giving a final score for teacher style dimensions (Benoit, 2015). The advantage of observation instruments is that they are used directly in the reality of the classroom instead of, like questionnaires, being a means for a more distant reflection of one's own teacher style. Thirdly, the instrument is user-friendly as – after training – the time needed to complete the observation procedure is limited to 10-15 minutes. Therefore, the instrument can not only be used by observers in the context of research studies, but also by teachers themselves in the reality of their own classroom, in order to improve their style.

Over the years, different versions of the ASOS have been developed, mainly to improve its user-friendliness (Benoit, 2015). The reliability and validity of the most recent version of the ASOS (the ASOS-E) in primary school contexts has not yet been thoroughly investigated. However, some studies have explored the reliability and validity of an earlier version of the ASOS: the ASOS-C.[54] In this version, after each observation period the heads were scored on a 6-point scale.

[54] The next overview concerning the results with regard to the validity and reliability of the ASOS-C was based on Benoit (2015).

Concerning validity, in the longitudinal study 'School careers in elementary education' (SiBO)[55], the internal structure of the ASOS was explored by conducting factor analyses (n = 50 schools, 4th and 5th grade) and calculating correlations between the subscales (now 'heads') and the global dimension scores (Goossens et al., 2009; Van Droogenbroeck et al., 2010). Factor analyses revealed three dimensions and 59,9%[56] of the collective variance was explained. The item 'Gives no clarity, provokes insecurity' vs. 'Gives clarity about situation and expectations' – called 'indicator' in the new instrument – loaded higher on Giving Autonomy than on Sensitivity. Correlations between the items[57] and the global dimension score on the 7-point scale ranged from $r = .78$ to $r = .81$[58] ($p < .01$). In a large piece of research conducted in Flanders called 'GOK'[59] (GOK, 2009) (n = 8 schools, 21 classes) it was found that the dimensions both with regard to the global scores as well as with regard to the mean item-scores were strongly correlated (ranging from $r = .65$ to $r = .90$, $p < .001$). Moreover, in this study it was concluded that more stimulation, more sensitivity and more support of autonomy by the teacher are fruitful for the observed engagement in pupils. This provided some evidence for the predictive validity of the ASOS.

In SiBO, the reliability of the ASOS-C also showed internal consistency ranging from .81 to .91[60]. However, in the study of Gouwy et al. (2002) a rather low interrater reliability was found. The authors concluded that there was an insufficient clarity about the content of the 'Giving Autonomy' dimension in comparison with the 'Stimulation' dimension and an intensive training was needed to prepare observers for scoring with the tool. In a new version, the ASOS-D, this problem was addressed by slightly redefining the dimensions and subdividing each head into different categories. Despite the fact that inter-rater reliability increased and less mental effort from observers was required (and therefore a less intensive training), the observation system was rather closed as every subdimension or head under the dimension had to be scored even if the number of observations relevant for it (critical incidents) was close to nil. Therefore, a new and final version, the ASOS-E – which builds further on the more open ASOS-C – has been developed. To achieve highly reliable coding and user-friendliness, in the more open scoring procedure of the ASOS-E the heads are seen as aspects which do not necessarily have to be scored when observing teacher style. After all, it is possible that the scorer would not observe interactions belonging to that particular head during the observation time. To further increase reliability, a manual is being developed in which each of the heads is

[55] The 'Schoolloopbanen in het basisonderwijs' (SiBO) research has tried to map the school careers of students, also by getting insight in the class context and what happens in class.
[56] Only based on the data for the 4th grade.
[57] Overall, the items were more or less the same as the indicators of the present instrument (see Figure 1).
[58] Only based on the data for the 4th grade.
[59] The 'Gelijke Onderwijskansen (GOK) research' is in-depth research set up in Flanders (Belgium) by 'Steunpunt Gelijke Onderwijskansen', in which it was investigated how primary school teachers deal with heterogeneous and underprivileged groups.
[60] Based on the data for the 4th and 5th grade.

thoroughly elaborated. Furthermore, short, intensive training sessions are being and will be organised to familiarise observers with the observation tool.

In the current study our aim is to check the validity, and more specifically the concurrent validity, of the most recent version of the ASOS (the ASOS-E). This form of validity has not previously been investigated for the ASOS. Concurrent or convergent validity is defined as the relationship between measures of the same construct using different assessment techniques (Crocker & Algina, 1986). In this study, we want to make a comparison with a well-known, frequently used instrument in educational research, namely the Classroom Assessment Scoring System (CLASS) (Pianta et al., 2012). In what follows, we will first describe the CLASS instrument and its theoretical background. Next, we will identify the parts which are congruent at the theoretical and operational level.

The Classroom Assessment Scoring System (CLASS) (Upper Elementary)

The CLASS (Pianta et al., 2012) measures the quality of classroom interactions among teachers and students. As such, Pianta et al. (2012) do not use the concept of teacher style, but conceptualise the concept of classroom interactions – not only between the teacher and students, but also among students – in 3 domains and 10 dimensions (see Table 1) which are similar to the 3 dimensions proposed in the experiential education theory. The framework, as defined by Pianta et al. (2012), also contains dimensions that are not fully part of teacher style but belong more to class climate as defined in experiential education theory. While the ASOS was built on one single theoretical framework, "the CLASS was developed based on an extensive literature review as well as on scales used in large-scale classroom observation studies" (Pianta et al., 2012, p.1).

The CLASS consists of the domains Emotional Support, Classroom Organisation and Instructional Support, and their dimensions (Pianta et al., 2012). It is noteworthy that the domains of the CLASS tool are the same across all grade levels, whereas the individual dimensions which are part of those domains vary to provide a context-specific and developmentally-sensitive metric for each age group. The dimensions found in Table 1 are for upper primary school contexts, as the convergent validity with the ASOS-E at this school level will be explored. The dimensions were derived after reviewing constructs assessed in classroom observation instruments, which have been used in educational research, focus groups, and extensive piloting.

Firstly, the emotional support domain encompasses interactions that reflect the emotional climate of the classroom. This is conceptualised as the warmth and/or negativity present in the classroom interactions, as well as the emotional connection between the teacher and the students [Positive Climate dimension]. Emotional support also includes a teacher's awareness of and responsiveness to students' levels of academic, social/emotional and developmental needs (NICHD ECCRN, 2002) [Teacher Sensitivity dimension]. Being emotionally supportive also requires that teachers take students' needs for relevant content, autonomy and interactions with peers into account [Regard for Student Perspectives dimension]. The dimensions of this domain are based on different research studies showing that these aspects are critical to school success (Allen, Hauser, Bell, & O'Connor, 1994; Allen, Kuperminc, Philliber, & Herre, 1994; Allen et al., 2002; Ryan & Deci, 2000).

Secondly, the classroom organisation domain includes the interactions with regard to managing time, behaviour, and attention in the classroom (Hamre & Pianta, 2007; Pianta et al., 2008). These interactions contain teachers' efforts to effectively manage class time [Productivity dimension], prevent and redirect misbehaviour (including misbehaviour directed to peers), and direct students' attention through clear and consistent organisation [Behaviour Management dimension]. This also means that unpredictable behavior like anger, hostility, aggression, or disrespect of teachers and/or students in the classroom does not fit in a good classroom organisation [Negative Climate dimension]. The theoretical basis of this domain can be found in the work of developmental psychologists studying self-regulatory skills (Blair, 2003; Raver, 2004), ecologist psychologists who study the extent to which different contexts contribute to these skills (Kounin, 1970, in Pianta et al., 2012), and in constructivist theories on engaging students in learning. "This work suggests that students develop better self-regulatory habits in well-regulated classroom environments" (Pianta et al., 2012, p.3).

Thirdly, the instructional support domain contains the quality of the instructional interactions between teachers and students with regard to the richness of the instruction and feedback provided [Quality of Feedback dimension] (Hamre & Pianta, 2005; Hamre & Pianta, 2007; Pianta et al., 2008). This domain has its main basis in research on students' cognitive and language development (e.g. Catts, Fey, Zhang, & Tomblin, 2001; Taylor, Pearson, Peterson, & Rodriguez, 2003), which highlights gaining 'usable knowledge' instead of simply learning facts [Content Understanding dimension]. Therefore, it is important that students learn how facts are interconnected, organised, and conditioned upon one another (Mayer, 2002). Next, to effectively stimulate the natural problem-solving abilities and curiosity of students, opportunities to solve ill-defined problems (Davidson & Sternberg, 2003), to apply learning to real world and novel contexts (Bransford et al., 2000, in Pianta et al., 2012), to utilise high-order thinking skills (Marzano, Pickering, & Pollock, 2001, in Pianta et al., 2012; Wenglinsky, 2002) and to stimulate metacognitive processes (Bransford et al., 2000, in Pianta et al., 2012; Pressley & El-Dinary, 1993) should be created [Analysis and Inquiry dimension]. While the Content Understanding and Analysis and Inquiry dimensions refer to the cognitive level(s) a teacher uses or to the expected levels of student responses, the Quality of Feedback dimension deals more with the amount of feedback provided (Pianta et al., 2012). Furthermore, an important starting point in this domain is that students learn more when they have opportunities for deep and meaningful conversations about content (Wolfs & Alexander, 2008, in Pianta et al., 2012) [Instructional Dialogue dimension]. Finally, variety and novelty in modes of presentation and types of activities (Cotton, 2000, in Pianta et al., 2012; Wenglinsky, 2000) are viewed as important for enhancing students' engagement [Instructional Learning Formats dimension].

'Student Engagement', an extra dimension which does not belong to one of the domains, was added. This dimension measures to what extent all students in the class are focused and participate in the learning activity presented or facilitated by the teacher. A distinction is made between passive and active engagement. According to the developers of CLASS, along with classroom and teacher aspects, it is also important to measure students' behaviour. Positive educational conditions can lead to positive student beliefs about their competence, changes in values and goals, and improved social connections (National Research Council, 2004, in Pianta et al., 2012). These can in turn lead to increased levels of academic engagement, motivation, and ultimately academic achievement.

The dimensions of the CLASS are further subdivided into behavioural indicators and markers, which should help in order to give a score for a particular CLASS dimension. The 'Instructional Dialogue' dimension, for example, is subdivided into three indicators: namely 'cumulative content-driven exchanges', 'distributed talk' and 'facilitation strategies'. The indicator 'distributed talk' consists of 4 behavioural markers: student-initiated dialogues, balance of teacher and student talk, majority of students and peer dialogues. Using the indicators and markers, a score is given for each dimension on a 7-point scale. The indicators are situated in the low, mid or high range, and the behavioural markers help to situate these indicators in one of these ranges.

Table 1. The CLASS Domains (Emotional Support, Classroom Organisation and Instructional Support) with their Dimensions.

Emotional Support	**Classroom Organisation**	**Instructional Support**
Positive Climate	Behaviour Management	Instructional Learning Formats
Teacher Sensitivity	Productivity	Content Understanding
Regard for Student Perspectives	Negative Climate	Analysis and Inquiry
		Quality of Feedback
		Instructional Dialogue
	Student Engagement[61]	

Note: More detailed descriptions of the domains and their dimensions can be found in Table 1 of Hafen et al. (2015)

The validity and reliability of the CLASS have been proven. Concerning validity, one can at first conclude that CLASS shows considerable face and construct validity (Pianta et al., 2012). While developing the instrument, numerous experts in classroom quality and teaching effectiveness have agreed that the CLASS measures aspects of the classroom that are of importance in determining student performance. Secondly, in the Measures of Effective Teaching (MET)[62] study, construct validity was determined by exploring the correlations with measures of related constructs (the FFT[63], UTOP[64], MQI[65], PLATO[66]). The correlations that were found are quite high (ranging from .68 to .88, $p < .001$), which suggests that the CLASS is capturing something similar in comparison to other instruments of quality teaching. Finally, evidence for predictive validity – namely the association of CLASS scores with students' performance – was provided. Results from the Secondary MyTeacher Partner Study show that classroom quality, as measured by CLASS, is associated with gains in children's performance in middle and high school classrooms (Allen et al., 2013). Also in the MET study, it was concluded that teachers who showed the types of practices emphasised in the CLASS had higher value-added scores than teachers who did not.

[61] The dimension 'Student engagement' is not integrated in this table. This dimension is an extra dimension not measured in the domains.
[62] www.metproject.org
[63] FFT = Framework for Teaching
[64] UTOP = UTEACH Observation Protocol
[65] MQI = Mathematical Quality of Instruction
[66] PLATO = Protocol for Language Arts Teaching Observation

With regard to reliability, evidence suggests that CLASS scores, assigned by trained, certified observers, are highly reliable (Pianta et al., 2012). Factor analyses and internal consistency estimates show that the dimensions of each domain are highly consistent characteristics of classrooms (alphas ranging from .91 to .92). Correlations between CLASS scores assigned in the autumn and spring showed to be low to moderate, indicating moderate stability over time (r between .26 and .49, $p < .001$). Data concerning inter-rater agreement show that observers consistently assign scores that are within one point on the scale. The results of four studies (Secondary MTP, MET and the Understanding Teaching Quality in Algebra Study (UTQ-A)) show that agreement within one point on the scale ranges from 64% to 98%, indicating moderate to high levels of agreement. Observers have an exact match (i.e. assign exactly the same code) about 30% of the time.

Theoretical relationship between the ASOS-E and the CLASS

In this study we want to determine the concurrent validity of the ASOS-E dimensions with relevant CLASS dimensions. In what follows, we provide our hypotheses concerning the theoretically similar dimensions. The formulation of these hypotheses is based on the description of the CLASS dimensions, and their constitutive indicators and behavioural markers[67]. As outlined above, no scores are given on individual indicators and markers when scoring with the CLASS[68]. Therefore, the concurrent validity analyses will be limited to the theoretically similar dimensions, even when only one indicator or marker is corresponding to what is measured in one or more of the ASOS-E dimensions. Table 2 provides an overview of the two instruments with their similar dimensions and indicators. This table functions as the basis for our analysis of the concurrent validity of the ASOS-E with the CLASS. To do the analysis, the CLASS manual was used (Pianta et al., 2012).

[67] When the description of the indicator or behavioural marker in itself does not make clear how the ASOS-E dimension relates to it, an example will be provided.
[68] While scoring, only the indicators are situated in the low, mid or high range.

Table 2. The ASOS-E Dimensions and their Corresponding CLASS Dimensions and Indicators.

	ASOS-E dimensions		
	Stimulation	***Sensitivity***	***Giving Autonomy***
Corresponding CLASS dimensions and indicators	Content Understanding ➔ *Depth of Understanding* ➔ *Communication of Concepts and Procedures* ➔ *Background Knowledge and Misconceptions* ➔ *Transmission of Content Knowledge and Procedures* ➔ *Opportunity for Practice of Content Knowledge and Procedures*	Teacher Sensitivity ➔ *Responsiveness to Academic and Social/Emotional Needs And Cues*	Regard for Student Perspectives ➔ *Flexibility and Student Focus* ➔ *Support for Autonomy and Leadership*
	Instructional Learning Formats ➔ *Learning Targets/Organisation* ➔ *Variety of Modalities, Strategies, and Materials* ➔ *Active Facilitation*	Positive Climate ➔ *Positive Communications* ➔ *Respect*	Analysis and Inquiry ➔ *Metacognition*
	Analysis and Inquiry ➔ *Opportunities for Novel Application* ➔ *Metacognition*	Negative Climate ➔ *Negative Affect* ➔ *Punitive Control* ➔ *Disrespect*	Quality of Feedback ➔ *Encouragement and Affirmation*
	Quality of Feedback ➔ *Feedback Loops* ➔ *Scaffolding* ➔ *Building on Student responses*	Quality of Feedback ➔ *Encouragement and Affirmation*	Instructional Dialogue ➔ *Distributed Talk*
	Instructional Dialogue ➔ *Cumulative Content-Driven Exchanges* ➔ *Distributed Talk* ➔ *Facilitation Strategies*	Productivity ➔ *Routines*	
	Regard for Student Perspectives ➔ *Connections to Current Life* ➔ *Meaningful Peer Interactions*	Behaviour Management ➔ *Clear Expectations*	
	Teacher Sensitivity ➔ *Responsiveness to Academic and Social/Emotional Needs and Cues*		
	Productivity ➔ *Maximising Learning Time* ➔ *Routines* ➔ *Transitions*		

Note: Only those indicators of the CLASS dimensions relevant for the ASOS(-E) dimensions are given

CLASS dimensions corresponding to the ASOS dimension Stimulation

Firstly, from a theoretical viewpoint, the ASOS dimension 'Stimulation' matches the CLASS dimension 'Content Understanding'. In 'Content Understanding', the extent to which "interactions among the teacher and students lead to an integrated understanding of facts, skills, concepts and principles" (Pianta et al., 2012, p.70) is measured. In particular, the indicator 'Depth of Understanding', which measures whether "the focus of the class is on encouraging deep understanding of content through the provision of meaningful, interactive discussion and explanation of broad organising ideas" (Pianta et al., 2012, p. 70), is related to 'Stimulation'. More specifically, this indicator relates to the stimulation of thinking (e.g. the teacher who stimulates a high depth of understanding explains how climate change has affected animal migration patterns; a teacher who does not stimulate this understanding asks, for example, to look up and memorise dictionary definitions for their vocabulary words (Pianta et al., 2012, p.78)). Content Understanding also has other indicators – 'Communication of Concepts and Procedures', 'Background Knowledge and Misconceptions', 'Transmission of Content Knowledge and Procedures' and 'Opportunity for Practice of Procedures and Skills' – that can be taken into account in the ASOS dimension Stimulation. These indicators assess whether the communication of concepts is effective (e.g. by highlighting the essential components, using examples); whether new concepts and ideas are connected to students' prior knowledge and misconceptions are clarified; whether knowledge is effectively and accurately communicated to students (with definitions, clarifications and rephrasing); and in case of procedures and skills taught in the lesson, whether students get opportunities for practice of procedures and skills (Pianta et al., 2012). The first three of these indicators relate to Providing Information; the last one to Stimulating Action.

Next, a connection between Stimulation (Stimulating Thinking/Stimulating Communication) and the 'Analysis and Inquiry' dimension can be found. The 'Analysis and Inquiry' dimension measures whether the teacher consistently scaffolds the processes in ways which allow students to be successful (through questions and support) while they solve new and/or open-ended problems, tasks, and questions ('Opportunities for Novel Application'), and whether students get opportunities to think about their own thinking through explanations, self-evaluations, reflection, and planning ('Metacognition'). The indicator 'Facilitation of Higher-Order Thinking' evaluated in this dimension corresponds less to the dimension Stimulation. With this indicator the observer can determine whether students get opportunities to engage in cognitively challenging tasks (e.g. investigating questions; examining, analysing or interpreting information or approaches; predicting; thinking about alternatives, and so on).

Convergent validity with 'Stimulation' (Stimulating Thinking, Stimulating on Communication, Stimulating Action) can also be determined for the 'Quality of Feedback' dimension that "assesses the degree to which feedback expands and extends learning and understanding" (Pianta et al., 2012, p.89). Similarities are found with three of its indicators: 'Feedback Loops', 'Scaffolding', 'Building on Student Responses' and 'Encouragement and Affirmation'. The first indicator evaluates whether there are frequent back-and-forth exchanges that lead to a deeper understanding of material and concepts; the second whether the teacher and/or peers scaffold students' learning (with assistance, hints, questions via which students are asked to explain their thinking, and so on); and the third whether the teacher and/or students expand, clarify or provide specific feedback on students' responses. The last indicator 'Encouragement and Affirmation' deals with recognising students' efforts and encouraging persistence.

Furthermore, a correspondence can be found with the dimension 'Instructional Dialogue'. As this dimension evaluates whether content-focused discussion among teachers and students is cumulative (indicator: 'Cumulative Content-Driven Exchanges'; behavioural marker: 'Depth of Exchanges') and whether there is distributed talk (indicator: 'Distributed Talk'; behavioural markers: 'Balance of Teacher and Student Talk' and 'Peer Dialogues') (Pianta et al., 2012), this dimension corresponds to the stimulation of communication in the ASOS. The first behavioural marker would be evaluated positively when the teacher asks the students to break into small groups to discuss which cause of the Civil War is the most important and why, for instance (Pianta et aL, 2012, p. 102); the second two, for example, when the teacher encourages students to share their ideas about why they emphatised with the main character (Pianta et al., 2012, p. 102) of a book or when students spontaneously start to whisper to one another about the book. There is also a connection between the dimension Instructional Dialogue and Stimulation Thinking in the ASOS. Instructional Dialogue is questioning whether teacher and students build on one another in discussions (behavioural marker: 'Exchanges that build on one another' of the indicator 'Cumulative content-driven exchanges') and whether they facilitate extended dialogues (e.g. by asking open-ended questions and statements, by listening actively...) (indicator: 'Facilitation Strategies'). Except for a few behavioural markers, all Instructional Dialogue indicators fit with the ASOS dimension Stimulation.

Moreover, a connection can also be found with the indicator 'Meaningful Peer Interactions' of the dimension Regard for Student Perspectives (behavioural marker: 'Peer Sharing and Group Work'). Similar to Stimulation Communication in the ASOS, this indicator assesses teachers' encouragement of students to talk to one other. The indicator 'Connections to current life' fits Providing Information in the Stimulation dimension in the ASOS. This dimension measures whether the teacher connects content to the world of students and explains the usefulness of specific content and skills. The other two indicators of Regard for Student Perspectives – 'Flexibility and Student Focus' and 'Support for Autonomy and Leadership' – cannot be connected to Stimulation in the ASOS. The first indicator evaluates whether students can share their ideas, and whether the teacher has an eye for students' responses and uses these responses in the lesson; with the second, whether students get meaningful choices in lessons and authentic opportunities for leadership and responsibility.

Furthermore, because of one behavioural marker – re-engagement – of the 'Responsiveness to Academic and Social/Emotional Needs and Cues' indicator, Stimulation (Stimulating Action) also converges to the dimension Teacher Sensitivity in the CLASS (e.g. the teacher notices a student who is no longer engaged (an academic need), walks over to him, and asks him a question to help focus his attention (Pianta et al., 2012). The other behavioural markers of that indicator, as well as the other indicators – 'Awareness', 'Effectiveness in Addressing Problems' and 'Student Comfort' – do not correspond to Stimulation. These indicators assess whether the teacher is aware of what students are doing, whether he/she is effective at helping students and whether students feel comfortable with the teacher.

Next, the indicator 'Active Facilitation' of the CLASS dimension 'Instructional Learning Formats' has two markers: 'Promoting Involvement' (e.g. after telling students about the topic for their journal writing, the teacher circulates around the room and stops to talk to students about their ideas instead of being uninvolved in what they are doing) and 'Teacher interest' (e.g. the teacher comes to class dressed as a soldier and shares reproduction artifacts that she has collected at different battle sites to help share her enthusiasm for the Civil War). These are similar to the heads 'Introducing Activities' and 'Providing Information' and 'Stimulating action' of the Stimulation dimension of the ASOS. The other marker of that indicator – as well as the indicator 'Effective Engagement' – do not theoretically fit with the ASOS dimension Stimulation. This indicator evaluates whether students appear engaged in the instruction (e.g. by volunteering, raising hands, participating). Also the indicators 'Learning Targets/Organisation' and 'Variety of Modalities, Strategies, and Materials' can be related to the Stimulation dimension (Providing Information/Stimulating Action). These indicators evaluate whether the teacher gives clear learning targets and presents information in an organised way; whether the teacher uses more than one modality (not only auditory) to present information, more than one strategy, and different materials which students have the opportunity to actively use.

Finally, Stimulation (Providing information) matches the indicators 'Routines' (behavioural markers: 'Students Know What to Do' and 'Clear Instructions'), 'Maximising Learning Time' and 'Transitions' of the Productivity dimension. The indicator 'Preparation' does not match the Stimulation dimension. These indicators measure whether students know what to do, whether the time for learning is maximised (e.g. by giving options for students who finish a task early), how smooth transitions occur and whether the teacher is fully prepared for activities (e.g. materials easily accessible).

CLASS dimensions which correspond to the ASOS dimension Sensitivity

Secondly, three CLASS dimensions measure similar aspects to the ASOS dimension 'Sensitivity'. The dimension 'Teacher Sensitivity' is in line with 'Sensitivity' (Understanding, Attention, Affection) as it assesses the teacher's timely responsiveness to the academic, social/emotional, behavioural and developmental needs of individual students and the entire class (e.g. when a teacher notices that a student has trouble with some problems – an academic need – he or she suggests to meet the next day during the planning period in order to figure them out together) (Pianta et al., 2012). While the indicator 'Responsiveness to academic and social/emotional needs and cues' relates to Sensitivity in the ASOS, the other indicators of the dimension Teacher Sensitivity – Awareness, Effectiveness in Addressing Problems and Student Comfort – are not directly evaluated in the ASOS dimension.

Next, the dimension 'Positive Climate' partly fits 'Sensitivity' (Respect, Affirmation), as this dimension evaluates whether "the teacher and students provide positive comments or indicate positive expectations" (Pianta et al., 2012, p. 21) and whether the teacher and students demonstrate respect (e.g. by using respectful language, students' names, or a warm, calm voice…). As well as the indicators 'Positive Communications' and 'Respect', Positive Climate has two other indicators – 'Relationships' and 'Positive Affect' – which are not connected to the ASOS dimension Sensitivity. The category Relationships assesses whether the teacher and students enjoy warm and supportive relationships with one another (Pianta et al., 2012, p. 24) and with Positive Affect whether teacher and students are enthusiastic.

Also, a less 'Negative Climate' is presumed/intended to correlate with a higher 'Sensitivity' (Respect, Affection, Clarity) as this dimension assesses the frequency, quality, and intensity of teacher and student negativity. The indicators 'Negative Affect', 'Punitive Control' and 'Disrespect' evaluate whether there is absence of strong negative affect, yelling at students or threatening to punish, and if the teacher and students are only very rarely sarcastic or disrespectful to one another.

One indicator of the 'Quality of Feedback' dimension (i.e. 'Encouragement and Affirmation') can also be connected to the Sensitivity dimension in the ASOS (Reinforcement). This is not the case for the indicators 'Feedback Loops', 'Scaffolding' and 'Building on Student Responses'.

Finally, the indicator 'Clear Expectations', in Behaviour Management (behavioural markers: 'Explicit', 'Consistent' and 'Students Know What to Do') has a connection with Sensitivity. This indicator evaluates whether behaviour expectations are clearly stated or understood by everyone in class. This does not apply for the indicators 'Proactive', 'Effective Redirection of Misbehaviour' or 'Student Behaviour', which determine whether the teacher prevents the development of problem behaviours, whether he/she uses effective strategies to redirect misbehaviour (which do not result in a loss of time); and whether students are compliant and only a few instances of student misbehaviour occur.

CLASS dimensions corresponding to the ASOS dimension Giving Autonomy

Thirdly, to the ASOS dimension 'Giving Autonomy' (Initiative, Method of Working, [final] product) fits the dimension 'Regard for Student Perspectives'. This measures how flexible the teacher is for students' leads, ideas and opinions and whether "students are provided with meaningful choices within lessons and are given authentic opportunities for leadership and responsibility" (Pianta et al., 2012, p.35). Two of the four indicators – 'Flexibility and Student Focus' and 'Support For Leaderships and Autonomy' – fit along with most of their markers to the ASOS dimension. The indicators 'Connections to Current Life' and 'Meaningful Peer Interactions' have no theoretical relation with 'Giving Autonomy'.

Next, the CLASS dimension 'Analysis and Inquiry' is related to 'Giving Autonomy' (Initiative) as the indicator Metacognition (or more specifically, the behavioural marker 'Students plan') evaluates whether students receive opportunities to plan (e.g. a teacher scoring high on this marker fully agrees and let students take the necessary steps when they approach him or her about doing a newspaper, instead of agreeing and then assigning other tasks to them) (Pianta et al., 2012). The other behavioural markers under this indicator and the other indicators of Analysis and Inquiry – 'Facilitation of Higher-Order Thinking' and 'Opportunities For Novel Application' – do not fit the dimension 'Giving Autonomy'.

Furthermore, the Quality of Feedback dimension is similar to the Giving Autonomy dimension (Initiative, Method of Working) because of one of its indicators: namely 'Encouragement and Affirmation'. This category evaluates whether students receive encouragement from the teacher which increases involvement and persistence for their efforts (Pianta et al., 2012) (e.g. when a teacher sees that a group of students is working very hard on their project, a teacher scoring low on 'Encouragement and Affirmation' may say that it is taking a lot longer than it should and that students should pick an easier topic next time). The other indicators of 'Quality of Feedback', 'Feedback Loops', 'Scaffolding' and 'Building on Student Responses', do not match to the Giving Autonomy dimension.

Finally, the indicator 'Distributed talk' – and more specifically two of its behavioural markers, 'Student-Initiated Dialogues' and 'Balance of Teacher and Student Talk' – of the dimension Instructional Dialogue correspond to 'Giving Autonomy' (Initiative). 'Student-Initiated Dialogues' evaluates whether the teacher builds further on thematic conversations initiated by students (e.g. when a student initiates a conversation in the classroom by bringing up the topic of food combinations that can be made in the lunchroom, the teacher asks students to brainstorm the food combinations and discuss which ones are the healthiest) (Pianta et al., 2012). The behavioural marker 'Balance of Teacher and Student Talk' measures to what extent students, relative to the teacher, take an active role (e.g. students discuss and create their own laws for the settlement of Jamestown while the teacher only occasionally prompts students' thinking). The other two markers of 'Distributed Talk', as well as the indicators 'Cumulative Content-Driven Exchanges' and 'Facilitation Strategies', do not connect to 'Giving Autonomy' in the ASOS.

Goals

The theoretical analysis allows us to conclude that although the CLASS dimensions measure other aspects not included in the ASOS-E, the CLASS measure is a good instrument to determine the concurrent validity of the ASOS-E. Just like the ASOS-E, it is an observation tool in which the interactions between teacher and students are measured, and which has various similar dimensions. Only the heads 'Rules and Agreements' and 'Conflict', under the ASOS dimension Autonomy, were found to have no corresponding CLASS dimensions.

Finding evidence for the concurrent validity of the ASOS-E would be advantageous for future research. The ASOS-E differs from the CLASS measure in that only teacher-student interactions are measured, while in the CLASS interactions among students are also taken into consideration. Furthermore, the scoring of the CLASS takes quite a long time – in principle 4 x 15 minutes observation during one lesson, apart from the scoring (Pianta et al., 2012) – and requires very intensive training from the observer, as 11 dimensions and 43 indicators have to be scored with this tool. Moreover, even after training, one needs to go through the main pages of the 11 dimensions in the CLASS manual when giving a score. For teachers and researchers, it would be more attractive to have a shorter, more compact instrument to evaluate teacher style.

Method

Participants

A total of 22 primary school teachers (19 female teachers and 3 male teachers) from 14 different schools, both Flemish and Dutch, participated in this study. The participant sample was drawn from a group of 34 teachers from the 4^{th}, 5^{th} and 6^{th} grade of primary school, who participated in a larger intervention study in which the effect of the implementation of a project-based science and technology learning environment called 'Village@School' was explored. In the next part we will explain how the sample was composed.

Data collection

As part of the data collection in the above-mentioned larger intervention study, observations were conducted during three science and technology activities. In two parallel activities, teachers organised respectively the 'Building a bridge' and 'Building a tower' activities with their pupils. In these activities, teachers were invited to organise a session in which their students had to construct a bridge and a tower out of paper strips, according to some minimal guidelines provided by the researchers. The teachers had also implemented a larger project, called Village@School. In this project students are challenged to build a miniature site on a standard plate (1,22m by 2,44m) with as many as possible working technological applications. The three contexts, in essence, are typical inquiry- and design-based learning environments. In these activities it is intended that students search for solutions to scientific and technological problems on their own and with their peers, while the teacher monitors these learning processes. Observations had a duration of approximately 45 minutes. An observation was conducted during every implementation of the 'Building a bridge' and 'Building a tower' activities. During the implementation of the Village@School project two observations were conducted.

All classroom observations were video recorded. Lessons were recorded with one camera with a wireless microphone for the teachers. This made it possible to capture teachers' interactions with pupils, which is necessary in group work settings in which pupils are talking with each other and interactions with the teacher are not always clearly audible. As well as interactions with the teacher, interactions among pupils were also recorded. To do so, the camera was positioned sideways in the classroom, in order to have a global view of the class; mostly in a fixed position, unless the teacher and most of the students went to another room (which was sometimes the case while working on the project Village@School). While observing, the observer regularly checked whether the camera was still working and recording properly.

Due to pragmatic constraints, some observations were initially scored live with the CLASS, while others were scored on the video. For the 'Building a bridge' activity all observations were scored on video; for the 'Building a tower' activity half of the observations were scored on the video, while the other half were scored live. While all Village@School observations were scored live in principle, for some observations this was not possible because teachers unexpectedly organised some time for clearing up, or the session did not take as long as previously agreed upon. These observations, or at least cycles of these observations (see further), were scored on the video. As far as possible, Village@School observations scored live were used (as planned; see further), to be able to score them afterwards on the video to calculate the ICC (instead of scoring for the first time on video). After the scoring of the observations in each setting separately ((1) 'Building a Bridge' activity, (2) Village@School project, (3) 'Building a tower' activity), a selection of teacher observations that could be re-scored on the video to calculate the ICC was made (as random as possible, but for the Village@School setting, also taking the observations which could be scored live on most of the dimensions). This has the consequence that the dataset consists of observations of the same teachers in more than one of the different settings which we provided. Of 8 teachers, two observations in different contexts (e.g. one during the 'Building a bridge' activity and one during Village@School) over the trajectory were integrated. A total of 30 observations could be used, spread over the three described settings. All of these observations were scored on the video using the ASOS.

Measures

The Adult Style Observation Schedule (ASOS-E)

To score with the ASOS, training is required. In the present intensive seminars[69] recently organised by the Centre for Experiential Education, training with the ASOS takes about approximately 235 minutes (I. Berghmans, personal communication, 31 mei 2016). This comprises a theoretical introduction, a profound analysis of a video clip and an assessment based on three video clips.

Two cycles of the recorded observation in the activities 'Building a bridge', 'Building a tower' and during the 'Village@School' project were coded on the basis of the video with the ASOS-E. Each observation cycle consisted of 10 minutes' observation. According to the ASOS guidelines, one should be able to give a

[69] The intensive seminars also contain training of other observation instruments, e.g. scoring well-being and involvement (Laevers et al., 2015).

score for the observation within a shorter period than is the case in the CLASS, for example. For each sample teacher, a general score for each dimension was calculated by averaging the available scores assigned in the two observation cycles. Half of the observations were scored by an Observer 1, whereas the other half of the observations was scored by an Observer 2.

To determine inter-rater reliability, the observations of 10 randomly selected teachers – with two cycles for each teacher – were scored by the two observers. An analysis was conducted by matching Observer 1's rating on each dimension of the ASOS to Observer 2's rating on each dimension. When two observers code the same cycle, they should consistently assign scores that are within one point on the scale (Pianta et al., 2012).

Tables 3 and 4 display the percentage of exact matches and exact and adjacent matches between the two observers for each dimension. When taking the scores for the separate cycles into consideration, the percentages of the exact plus adjacent matches for Stimulation, Sensitivity and Giving Autonomy were 80%, 100% and 90% respectively. For Stimulation, the strength in reliability is somewhat lower but still rather good, as is shown by the percentage of exact and adjacent matches for the mean scores of the two cycles per session. This lower strength may be due to the rather small sample of 10 teachers for which inter-rater agreement was calculated.

Together with an experienced expert researcher in the field Observer 2, who was less experienced with the ASOS-E, re-watched and discussed the cycles of a first few observations with a discrepancy higher than one. The opportunity to discuss the teacher style of a few lessons provided a vehicle to establish firmly her understanding of the signals for each dimension before continuing with the scoring of half of the observations. As well as this, other, new cycles from another teacher sample were scored in order to develop the new, more 'profound' understanding of Observer 2. For the cycles of the teachers with a discrepancy higher than one, a norm score was determined. This score came into existence after discussing the discrepancy between scores. This norm score was then used for the calculation of the mean of the two cycles. For those cycles for which the scores between the observers differed no more than one, the scores of Observer 1, and not of Observer 2, were used to calculate the mean over the two cycles.

Table 3. Double Coding Reliability for ASOS-E (cycles)

Dimension	% Exact	% Exact + Adjacent
Stimulation	40	80
Sensitivity	65	100
Giving Autonomy	40	90

Table 4. Double Coding Reliability for ASOS-E (mean of 2 cycles)

Dimension	% Exact	% Exact + Adjacent
Stimulation	20	70
Sensitivity	60	100
Giving Autonomy	20	80

The Classroom Assessment Scoring System (CLASS) Upper Elementary

The CLASS (Pianta et al., 2012) was used to measure the quality of classroom interactions among teachers and students. CLASS consists of 3 domains and 12 dimensions[70] (see Table 1).

The observer (first author) rated the observation on each dimension and indicated the quality of the interactions on a 7-point Likert scale ranging from 0 to 7. An intensive training of two days was completed beforehand in order to be able to score with CLASS. This is required in order to become a reliable observer (http://teachstone.com/services/training/class-observation-training-programs). After the training observers get 8 weeks of time to do a test in which five new video clips have to be scored using the CLASS measure. The observer needed one week to study the manual, practising the scoring of all dimensions with three video clips (about half a day to watch and score one clip) and another week to do the test. For the test it is recommended to spread the scoring of the five video clips over time in order to be objective as possible for each new observation. The observer obtained a certificate and passed an annual renewal test. Evidence suggests that CLASS scores, assigned by trained, certified observers, are highly reliable (Pianta et al., 2012).

Two observation cycles were conducted during the 'Building a bridge' and 'Building a tower' activities, and in each 'Village@School' session. In the case that the observation was initially scored live with the CLASS, an interval time of approximately 10 minutes was needed to actually score the observation.

[70] The dimension 'Student engagement' is not integrated in this table. This dimension is an extra dimension not measured in the domains.

When the observation was initially scored on the video (e.g. for the 'Building a bridge' activity), after the first 15 minutes of observation, the video could be stopped, and the immediate next 15 minutes were scored. As in the ASOS, one general score for each dimension was calculated, by averaging the scores assigned in each observation cycle. In CLASS it is recommended to score 4 cycles for one teacher to reliably draw conclusions (Pianta et al., 2012). Because of the limited duration of the session, it was not possible to observe 4 cycles of each activity in our fairly large sample of observations. As the sample is larger than in other studies in which observation instruments were used, we also expect to obtain reliable results.
In two other studies, we had already re-scored 38 observations on the video[71] over the three activities (the two parallel activities, and the two observations during Village@School) and calculated the ICC to ensure intra-rater reliability[72]. The 30 observations that constitute the data for the present study are part of these 38 observations. Following the criteria from Cichetti and Sparrow (1981), the ICC for the dimensions was fair to excellent, as ICCs ranged from .42 to .90; except for the dimension 'Instructional Learning Formats', which had an ICC of .26[73]. In other studies using CLASS, ICCs (two observers) ranging from .15 to .43 are reported (Hafen et al., 2015). This is not too problematic as these lower ICCs can be explained by the fact that in CLASS, adjacent agreement (scoring within one point) is allowed.

As all of the observations in this study were scored on video with the CLASS and the ASOS, the correlations with the video CLASS scores will be calculated. As not all of the cycles scored with the CLASS and the ASOS are exactly the same, correlations were only calculated for the mean of two cycles, but not for each cycle separately.

[71] For the observations scored on video in order to calculate the ICC, the initially scored (live or video) cycles were used.
[72] Because of the high cost of having the data double-scored by another trained and certified researcher, it was not possible to calculate inter-rater reliability. However, after having followed training, being certified and re-certified, one should be able to reliably score classroom interactions.
[73] Therefore, analyses with this dimension were excluded.

Results

In Tables 5, 6 and 7 descriptive statistics and correlations are presented.

Table 5 Means, Standard Deviations, Range, Skewness and Kurtosis of Study Variables (CLASS-domains during Village@School as outcome)

	M	SD	Range	Skewness (SE)	Kurtosis (SE)
1. Stimulation	4.78	1.01	4.50	-.25(.43)	.04(.83)
2. Sensitivity	4.68	.97	4.00	-.23(.43)	-.43(.83)
3. Giving Autonomy	4.46	.73	3.00	.13(.43)	-.51(.83)
4. Positive Climate	5.02	.65	2.50	-.67(.43)	-.39(.83)
5. Teacher Sensitivity	5.33	.58	2.50	-1.01(.43)	2.00(.83)
6. Regard for Student Perspectives	4.55	.50	2.00	-.44(.43)	-.40(.83)
7. Behaviour Management	5.57	.64	2.50	-1.85(.43)	3.36(.83)
8. Productivity	5.32	.38	1.50	-.24(.43)	-.04(.83)
9. Negative Climate (Rev.)	6.55	.63	3.00	-2.48(.43)	8.29(.83)
10. Content Understanding	4.05	.94	4.00	-.32(.43)	-.37(.83)
11. Analysis and Inquiry	4.55	.75	3.00	-.71(.43)	.47(.83)
12. Quality of Feedback	4.37	.74	3.00	.09(.43)	-.41(.83)
13. Instructional Dialogue	4.68	.69	2.50	-.04(.43)	-.85(.83)

Table 6 Bivariate correlations between the ASOS-dimensions

	1	2	3
1. Stimulation	1		
2. Sensitivity	.78**	1	
3. Giving Autonomy	.75**	.66**	1

Table 7 Bivariate Correlations between the CLASS-dimensions

	1	2	3	4	5	6	7	8	9	10
1. Positive Climate	1									
2. Teacher Sensitivity	.38*	1								
3. Regard for Student Perspectives	.13	.18	1							
4. Behaviour Management	.16	.31	.26	1						
5. Productivity	.15	.48**	.01	.48**	1					
6.Negative Climate (Rev.)	.35	.17	.07	.23	.40*	1				
7. Content Understanding	.28	.16	-.13	.48**	.43*	.20	1			
8. Analysis and Inquiry	.28	.22	.16	.48**	.24	.10	.62**	1		
9.Quality of Feedback	.38*	.13	-.05	.33	.09	.09	.73**	.78**	1	
10. Instructional Dialogue	-.01	.04	.25	.27	.13	.06	.35	.54**	.47**	1

Pearson product-moment correlation coefficients between the ASOS dimensions and CLASS domains/dimensions are presented in Table 8. Taken together, three of the 16 hypothesised correlations were significant, and in the expected direction. The magnitude of the correlations varied from -.02 to .48. Using the scale developed by Cohen (1988), correlations of .10 to .29, .30 to .49, and .50 and above were considered small, moderate, and large respectively. In order for a correlation to be cited as evidence of validity, it should demonstrate statistical significance (Anastasi & Urbina, 1996). Because of the small sample size, also correlations with a significance lower than .10 will be reported.

Table 8. Correlations between the ASOS dimension scores and CLASS dimension scores (mean of 2 cycles) (n = 30)

CLASS dimensions	**ASOS-E dimensions**		
	Stimulation	Sensitivity	Giving Autonomy
Positive Climate	.47**	**.48****	.33★
Teacher Sensitivity	**.12**[a]	**.12**	.05
Regard for Student Perspectives	**-.02**	-.03	**.17**
Behaviour Management	.28	**.06**	.20
Productivity	**.20**	**.08**	.06
Negative Climate (reversed)	.18	**.25**	-.01
Content Understanding	**.37***	.15	.07
Analysis and Inquiry	**.11**	**.07**	**-.01**
Quality of Feedback	**.33★**	**.27**	**.13**
Instructional Dialogue	**.12**	-.04	**.08**

Note: The correlations are calculated with the CLASS scores obtained via video observations.

[a]The correlations in bold are the correlations between dimensions which were theoretically expected to correlate with each other.

★ $p < .10$, *$p < .05$ and ** $p < .01$

For each ASOS-E dimension we expected correlations with different CLASS dimensions. In our theoretical analysis, in which both instruments were compared, we did not always find a whole CLASS dimension to correspond to an ASOS-E dimension, but one or more indicators or behavioural markers. In Table 2, the expected relations are given. When presenting the results, this table functions as the starting point to discuss the found correlations in Table 5.

Firstly, for the ASOS dimension Stimulation, two significant correlations were found; in particular with Positive Climate ($r = .47^{**}$, $p < .01$) and with Content Understanding ($r =.37^{*}$, $p < .05$). The positive relation with Positive Climate is unexpected. Differing from our hypotheses, no correlations were found with the dimensions Analysis and Inquiry, Quality of Feedback, Instructional Dialogue, Regard for Student Perspectives or Teacher Sensitivity. The correlation with Quality of Feedback is moderate but not significant at the .05 level.

Secondly, for the ASOS dimension Sensitivity, a significant correlation was found with the dimension Positive Climate ($r = .48^{**}$, $p < .01$). The expected correlations with Teacher Sensitivity, Negative Climate (reversed), Quality of Feedback, Productivity and Behaviour Management were not found. For Negative Climate (reversed) and Quality of Feedback, slightly higher – in comparison to most of the other dimensions – but still small and non-significant correlations were found for Sensitivity.

Finally, for the ASOS dimension Giving Autonomy, no significant correlations were found with the CLASS dimensions. We had hypothesised that they would correspond to the Giving Autonomy dimension (Regard for Student Perspectives, Quality of Feedback, Analysis and Inquiry and Instructional Dialogue). Though, the correlation with Positive Climate is significant at the .10 level.

Conclusion and discussion

Before now, the instruments that measure the quality of learning environments in primary school, via observation, were limited. In this study, we have explored the quality – in particular the concurrent validity – of the Adult Style Observation Schedule (ASOS), an instrument which aims to measure teacher style. Teacher style is conceptualised as the pattern of teacher interventions during classroom activities (Laevers, 2005; Pettigrew et al., 2013) and becomes visible in the interactions between teachers and their students (de Kruif et al., 2000; Laevers & Heylen, 2013). The ASOS dimensions were compared with the dimensions of an instrument which measures the quality of classroom interactions – namely the Classroom Assessment Scoring System (CLASS Upper Elementary) (Pianta et al., 2012) – and that, from a theoretical viewpoint, showed congruencies with the ASOS. While this instrument has its value, a lot of different dimensions have to be evaluated, which is time-consuming when investigating a large sample of teachers. The ASOS (Benoit, 2015) consists of only three dimensions and is therefore a more compact instrument. The ASOS is also more unambiguous as only interactions between the teacher and the students are scored.

The CLASS and ASOS scores used for the analyses were all obtained via video observations. We looked at the correlations for the mean of 2 cycles. In the CLASS, a teacher can only get a reliable score when it is composed of the mean of the scores on different cycles, and ideally on the basis of four cycles (Pianta et al., 2012). The concurrent validity analyses were based on a theoretical analysis in which the congruencies between the ASOS-E and the CLASS were explored. The correlation between scores obtained via the ASOS-E was correlated with scores obtained via the CLASS measure.

In the Stimulation dimension in the ASOS, different aspects are measured, of which some may be more decisive than others in making a general estimation of the Stimulation dimension. Maybe the most important signal in the ASOS deals with the stimulation of a deep level of thinking. Typical for the ASOS is that the teacher takes the perspective of the child (Benoit, 2015; Laevers & Heylen, 2003), determines whether the child is engaged in certain activities and materials – and tries to intervene in order to increase this engagement. This is done by giving substantial appealing information to learners, but happens more frequently by asking learners open-ended questions without immediately giving the right answer. As discussed, we remark that for the dimensions of the ASOS the concept of 'modus' is used to point to the potential of the interventions to increase (well-being and) engagement (Benoit, 2015; Lento, 2016). While in the CLASS the relation of different dimensions, indicators and markers with pupils' engagement is not made or not presumed, in the dimension Content Understanding – which was showed to correlate with Stimulation in the ASOS – this is more the case. In that dimension, the meaningfulness of the content presented is emphasised. Content can be made more meaningful by helping students to apply their thinking to real world events and situations and by presenting and probing multiple, varied perspectives. "The teacher is primarily concerned with students being able to understand the different perspectives and to support/substantiate whatever position they choose" (Pianta et al., 2012, p. 76). This is not so much the case for the CLASS dimension Quality of Feedback, in which the amount of feedback is scored, rather than whether this feedback really makes sense for students. In the CLASS manual one can read that "the focus here should be on the nature of the feedback provided and the extent to which it 'pushes' learning" (Pianta et al., 2012, p. 89). In the CLASS dimension Analysis and Inquiry – for which we also expected but did not find a relation with Stimulation in the ASOS – it is evaluated whether the teacher guides students' processes of working on novel problems and open-ended tasks, but not that this guidance should be related to students' needs. More than this, the Analysis and Inquiry dimension evaluates whether a teacher asks questions to students in order to make them reflect on or explain their own cognitive processes (indicator 'Metacognition'). However, it is possible for a teacher to receive a high score on this indicator even when the questions asked by the teacher are too academic or not tuned to the level of the average student. Such questions would however not reflect a high level of stimulation of thinking by the teacher in the ASOS-E. The Instructional Dialogue dimension, similar to the Content Understanding dimension, also evaluates whether dialogues are connected to content in order to further students' understanding. However, a correlation with Stimulation may not have been found as in the Instructional Dialogue dimension, the meaningfulness of this content is less emphasized than in the Content Understanding dimension.

Furthermore, while in Teacher Sensitivity and Regard for Student Perspectives some indicators deal with the re-engagement of students and the connection of content to their world, there are other, different indicators and behavioural markers evaluated in these dimensions which may have resulted in not finding the hypothesised correlations of Stimulation. The relation of Stimulation with Positive Climate was surprising with regard to our hypotheses. A possible explanation can be found in the modus, in the idea that the more enthusiastic you are as a teacher, the higher your scores for Stimulation. The signals of the ASOS show that activities should be introduced in a motivating manner, information should be presented in a captivating way and that the teacher invites communication in a stimulating manner. Not only learners, but also the teacher needs to be engaged and 'in flow' (Csikscentmihalyi, 1988). In the dimension Positive Climate, this teacher and student enthusiasm is measured via the indicator Positive Affect (with 'Smiling', 'Laughter' and 'Enthusiasm' as behavioural markers). A high score on this indicator means that "there are frequent genuine displays of positive affect among the teacher and students" (Pianta et al., 2012, p. 21).

Finally, in the open-ended S&T learning environments that we observed, a lot of interactions among pupils also took place. We could observe that, for example, a high score on Quality of Feedback could not automatically be contributed to interactions with the teacher due to the frequency of back-and-forth exchanges between pupils. While in principle the CLASS is designed to conduct observations in various contexts, in the manual some dimensions are operationalised in such a way that they give the impression of starting from whole class teaching, alternated with group activities. However, in whole class teaching the teacher has a more prominent role and interactions could be more clearly observed in comparison to more open-ended learning environments. The starting point in the ASOS is different as this instrument is developed in the context of experiential pre-school education – and developed further for primary and secondary education – with a strong emphasis on child-initiated activity.

With regard to the ASOS dimension Sensitivity, it is somewhat counterintuitive that no correlation was found with the CLASS dimension Teacher Sensitivity. In the theoretical analysis preceding the correlational analyses, it was however found that only one indicator of Teacher Sensitivity – Responsiveness To Academic And Social/Emotional Needs And Cues – corresponded theoretically to Sensitivity in the ASOS. Moreover, not all behavioural markers of this indicator apply to Sensitivity as measured in the ASOS, as some of them relate to responsiveness to students' academic needs ('Individualized Support', 'Reassurance and Assistance', 'Adjusts Pacing/Wait Time as Needed', 'Re-Engagement' and 'Timely Response'). In short, both dimensions have the same fundamental grounding, but in Teacher Sensitivity three other indicators which have no direct connection with Sensitivity as measured in the ASOS – 'Awareness', 'Effectiveness in Addressing Problems' and 'Student Comfort' – are also evaluated. Awareness, however, has to do with a teacher taking perspective, which is a precondition in each of the heads of the ASOS, but there are no aspects in the ASOS that explicitly evaluate that indicator. The correlation of the ASOS dimension Sensitivity with Positive Climate was expected as two indicators of Positive Climate – 'Positive Communications' and 'Respect' – directly correspond to Sensitivity. As it is also evaluated whether the teacher discourages students in the ASOS dimension Sensitivity, the slightly higher but not significant correlation with Quality of Feedback – the indicator 'Encouragement and Affirmation' was theoretically congruent – can be explained. Similarly, via its indicators, Negative Climate also measures affection and respect, which are also measured in the Sensitivity dimension of the ASOS.

Thirdly and surprisingly, no correlations were found between the Giving Autonomy dimension of the ASOS and the hypothesised corresponding CLASS dimensions Regard for Student Perspectives, Quality of Feedback and Instructional Dialogue. It is important to note that Regard for Student Perspectives and Instructional Dialogue contain indicators on which the teachers who participated in this study often scored 'automatically' higher as we provided them with open-ended S&T learning environments. For Regard for Student Perspectives this is the case for the indicators 'Connections to Current Life' (the Village@School project starts from students' own interests and realities) and 'Meaningful peer interactions' (almost all teachers organised the provided activities by giving the opportunity for students to work together). During the provided activities, in most classes class dialogues were distributed such that both the teacher and the majority of students took an active role or students were actively engaged in instructional dialogues with each other (Instructional Dialogue indicator 'Distributed Talk'). Taking these facts into consideration, two questions can be raised: (1) Is the CLASS, with each of its dimensions and indicators, able to distinguish teachers with a higher quality of interactions from teachers with a lower quality in open-ended learning environments in which a lot of initiative is provided to students? and (2) Does the CLASS only measure the quality of the interactions or also the quality of the provided learning environment?

Limitations of this study should be noted. Firstly, although the sample of 30 observations in teachers was – in light of the intensive data collection via observation – quite large, a larger sample may have resulted in higher correlations. The sample was also too small to conduct a factor analysis, which is usually done when exploring instrument validity. Secondly, the teachers involved in this sample participated on a voluntary basis to the larger intervention study. This fact, together with the hypothesis that the CLASS may not distinguish enough between teachers in open-ended learning environments, may have the consequence that the variation among teachers is rather small. It is possible that the results would be more congruent with our hypothesised correlations if repeated with a sample that was more evenly distributed, with teachers scoring extremely high or low on the dimensions of both CLASS and ASOS.

Future research could address the previous results and limitations in several ways. With a larger sample size, a more random selection of teachers and the observation of interactions/teacher style in less open learning environments, one could consider firstly whether it is desirable to further explore why some aspects measured in the CLASS are categorised in other dimensions than in the ASOS. Why is it, for example, that 'Positive Communications' and 'Respect' can be found under the dimension 'Positive Climate' and not under 'Teacher Sensitivity'? Or that a behavioural marker like 'Re-Engaging' (under the indicator 'Responsiveness to Students' Academic and Social/Emotional Needs') can be found under Teacher Sensitivity, while it also makes sense – like in the ASOS – that it is part of a stimulating teacher style?

Secondly, it would be worth to investigate whether the different dimensions, indicators and behavioural markers of the CLASS have the potential to realise engagement in students. A profound qualitative analysis on the level of the indicators and markers can provide insight into the dimensions that are likely to correspond to the ASOS-dimensions. The results of this analysis may further explain why different expected correlations were not found, and to which extent the embeddedness of the modus of the interventions in the direction of engagement makes the ASOS different from other tools that aim to measure teacher style.

Thirdly, some dimensions or indicators in the CLASS appear not to be limited to the interactions between teachers and students or among students, and thus to teacher style, but also to more 'robust' characteristics of the learning environment. This reflection does not only relate to an assumed side effect of the learning environment in which the observation is conducted (see above, e.g. little variation in Regard for Student Perspectives in an open-ended learning environment), but is also related to a theoretical analysis of the CLASS dimensions themselves. This applies, for example, to the Positive Climate in the classroom – the distinction from Negative Climate is, on a side note, not crystal clear – with an indicator like 'Relationships' which measures whether there are indications that the teacher and students enjoy warm and supportive relationships with one another (Pianta et al., 2012). Different dimensions measure whether the materials and activities provided are satisfying. This is the case for the dimension Instructional Learning Formats in which the indicator 'Variety of Modalities, Strategies, and Materials' measures whether the teacher uses a variety of modalities (e.g. auditory, movement, visual, or nonverbal expression and behaviour), strategies (e.g. small group discussions, writing, drawing), and (interactive) materials (with which students have the opportunity to manipulate or explore any resources or materials). The same applies to the dimension 'Content Understanding', and more specifically to the indicator 'Communication of Concepts and Procedures', which evaluates whether class discussion and materials consistently and effectively communicate the essential attributes of concepts and procedures to students (Pianta et al., 2012). As such, the concept of interactions appears to cover more than the particular communications between the teacher and students alone. However, the manual prescribes that the presence of materials, the physical environment or safety, and the adoption of a specific curriculum are not included in the CLASS scales (Pianta et al., 2012). At the Centre for Experiential Education, other instruments were developed to measure the climate and other more 'fixed' characteristics of the learning environment (e.g. whether the materials provided are realistic). In total, seven factors were distinguished (Laevers & Heylen, 2013). Though, it should be noticed that drawing a line between teacher style and these rather 'fixed' characteristics of the learning environment is a challenge. A teacher who introduces new materials in the course of the lesson, may be evaluated on his/her teacher style instead of on his/her initial offer of materials (for instance evaluated in 'Variety of Modalities, Strategies and Materials').

Fourthly, some dimensions (or their indicators) tend to measure the interactions between teachers and students, while others also aim to evaluate the interactions among students and still others seem to measure only the interactions among students. With CLASS, it is not only evident that teacher and student interactions are measured as well as interactions among students. It is also that on one occasion 'the teacher and students' (e.g. Positive Climate) is used in the description of the different scale values, on another occasion 'the teacher and/or students' (e.g. in most of the indicators of Quality of Feedback), and on still another occasion 'the teacher' (e.g. in two indicators of Regard for Student Perspectives). The rationale behind this choice and the ambiguity that exists for the observer has to be made clear, in order to gain insight into what is seen more as part of the interactions with the teacher alone and what is part of the interactions with students. Perhaps it is only meaningful to calculate correlations of the ASOS dimensions with dimensions of the CLASS which exclusively measure the quality of the interactions between the teacher and the students. An alternative would be to conduct the validity analyses with another observation instrument that solely focuses on teacher style.

Fifthly, there is a possible limitation of not being able to distinguish expert from less expert teachers when using the CLASS in open-ended S&T learning environments. To address this, validity analyses should be conducted in less open learning environments in which language skills are taught, for example. In a later phase it would be interesting to determine whether – in comparison to the CLASS – the ASOS is a more suitable instrument for the evaluation of teacher style in open-ended learning environments.

Finally, it would be interesting to check for other forms of validity. Predictive validity analyses could be used to investigate whether teacher style as measured by the ASOS contribute(s) to learners' affective and cognitive learning outcomes (engagement and development of competences), versus the classroom interactions as evaluated in the CLASS. Construct validity analyses can be conducted by asking experts in the field how excellent teachers can be discerned from less excellent ones when it comes to teacher style.

With this study, a step has been taken to optimise existing and develop new instruments for measuring a teacher's style. It is important to invest in the development of high-quality, user-friendly instruments which measure diverse aspects of how a teacher interacts with learners in order to gain more insight into his or her effectiveness in specific learning environments. The foundation for doing that has already been laid. In the literature there is a consensus about (1) the importance of the quality of interactions between teachers and learners, and (2) the characterisation of these interactions or the teacher style with stimulating, sensitive, and autonomy-supportive interventions. Having more knowledge about what such an effective role encompasses may give us insight into how teachers could be supported to strengthen the dimension which strongly determines the power of the learning environment.

Conclusions and discussion

The aim of this doctoral dissertation was twofold. The first goal was to investigate the existing literature with regard to instruments that assess the quality of project-based S&T learning environments as well as to contribute to instrument development. The second goal was to assess the effectiveness of the implementation of a project-based S&T learning environment on pupils' growth in engagement and on the growth of teachers' competence profile (i.e. their attitudes towards S&T (teaching) and their teacher style). In addition, the predetermining factors for a high-quality teacher style in the project-based learning environment, as well as the explaining factors for pupils' growth in engagement, were explored. In this general discussion, the main findings with respect to these two aims, which unfolded over four studies, are summarised and discussed. Moreover, a reflection on the limitations of the different studies is provided. Directions for future research, related to these limitations, are also incorporated in this final section. Finally, implications for educational practice are suggested.

Main findings

Research aim 1 – Reviewing the existing literature with regard to instruments in S&T and contributing to instrument development

Study 1 contributes to the field of S&T education research firstly, by providing a conceptual analysis of particular aspects of inquiry- and design-based learning environments which are measured by the available instruments in the literature. Secondly, this study reviewed not only questionnaires that measure pupils' perceptions, but also observation instruments, interview protocols and logbooks which focused on a broad range of aspects of S&T learning environments. Previous review studies (Fraser, 2012; Liu, 2010, 2012; Wubbels & Brekelmans, 2012) did not integrate all of these types of instruments when conducting their review of existing instruments which evaluate the quality of S&T learning environments.

Although different kinds of instruments exist in the field of S&T learning environments, it was until now difficult for researchers in the field to make a deliberate choice about which instruments, scales, items or questions to use. This review helps researchers who want to measure one or more particular aspects of the S&T learning environment. It encourages them to compose new instruments with scales, when operationalised, best fit best with their goals. Using instruments to empirically measure the quality of project-based S&T learning environments, or of S&T learning environments in general, is important to gain more insight into what works and what does not. This need is especially high when it comes to the role of the teacher in these learning environments (Polman & Pea, 2001).

One can also go a step further by selecting existing valid and reliable instruments which are used in fields other than S&T, but which measure other aspects of learning environments also mentioned in the literature on S&T. As an instrument which comprehensively evaluates the teacher's role was not found, the CLASS Upper Elementary (Pianta et al., 2012) – a widely recognised valid and reliable instrument also used in other educational fields – was selected. In light of the research tradition of CEGO and the aims of this dissertation, it was a conscious choice to use an observation instrument rather than a questionnaire, as is often done in other studies investigating open-ended S&T learning environments (e.g., Mant et al., 2007). However, there was a strong need to obtain profound insight into how teachers actually realise these learning environments (Krajcik et al., 1998); therefore, it was essential to use validated observation instruments that register what happens in the classroom.

After using the CLASS instrument in Studies 2 and 3, the opportunity arose to explore the qualities of an existing instrument developed at CEGO: the ASOS (Laevers & Heylen, 2013), created to assess the teacher's style in a wide variety of situations (i.e. with different contexts and content). In the fourth study, the concurrent validity of the ASOS with the CLASS was explored. The further improvement of the ASOS may not only contribute to the development of instruments in general – which can also be used in the fields of S&T – but may also open a new theoretical debate about existing instruments which measure teacher style or teacher-learner interactions. In Study 4 the advantages of the ASOS, in comparison to the CLASS, were discussed. Its user-friendliness and sole focus on the interactions between the teacher and the pupils was highlighted. The results of the validity analyses showed that the relationships of the CLASS dimensions with the ASOS dimensions that we expected – on the basis of a profound theoretical analysis of these dimensions and their operationalisation – were not univocally found. In particular, the ASOS dimension Stimulation showed congruence with the CLASS dimension Content Understanding, and the ASOS dimension Sensitivity correlated with the CLASS dimension Positive Climate. When discussing the results of Study 4, questions were raised with regard to the suitability of the CLASS in measuring teacher style in open-ended S&T learning environments. Some dimensions – such as Regard for Student Perspectives and Analysis and Inquiry – may not sufficiently distinguish between teachers in these settings. Moreover, while observing, it became clear that interactions among pupils may compensate for low-quality teacher-pupil interactions, with the result that a high score for that class observation is still obtained. As a consequence, the suggestion was made for future research to only calculate correlations with those CLASS dimensions which solely evaluate the teacher-pupil interactions. Finally, the ASOS sets the bar somewhat higher in comparison to the CLASS, as teacher interventions are evaluated more in light of their potential effectiveness for pupils' engagement. In addition, in the experiential education framework, more substantial requirements are set regarding pupils' engagement (Laevers et al., 2011). Simply stated, a pupil who participates actively in CLASS, is not necessarily engaged according to the experiential theory. Concerning the connection of interactions with pupils' engagement, a teacher can, for example, score highly on his/her feedback loops with pupils (characterised by frequent back-and-forth exchanges), but the nature of the questions asked may be not stimulating enough to increase pupils' engagement. From this perspective, the ASOS goes a step further in the assessment of teacher-pupil interactions.

Research aim 2 – Exploring the implementation of the project-based S&T learning environment Village@School: Effectiveness and (potential) determining factors

The second aim of the dissertation was to investigate the effects of implementing the challenging project-based S&T learning environment Village@School on pupils' possible growth in engagement on the one hand (Study 2) and on the growth in teachers' competence profile on the other (Study 3). To gain more insight into the role of teachers' competence profiles when implementing such a learning environment, the relation between teachers' attitudes towards S&T (teaching) and teacher style, both measured before the project, and the teacher style as evaluated during the project, was also investigated (Study 3).

In contrast to most studies which have focused on the effect of the implementation of particular S&T learning environments as a whole (e.g., Kaldi et al., 2011), the second and third study focused on the role of the teacher implementing the project. Such an in-depth study of the teacher's role not only addresses the ambiguity of the literature regarding this role (e.g., Polman & Pea, 2001), but also informs educational practice, and therefore has several merits. When we know what works and what does not, specialised teacher training can be organised, which is likely to have more visible effects. In the end, such training may make it easier for teachers to implement S&T learning environments in their classes.

On the pupil side, the investigation of the effectiveness of the project-based learning environment Village@School shed light on engagement. More specifically, there was a focus on pupils' growth in engagement. Pupils' evolution in engagement deserves special attention as most pupils are not used to learning environments in which they receive ample room for initiative. Pupils have to learn to make their way in these learning environments, as waiting passively for teacher instructions is no longer possible. Therefore, the question of how the teacher could contribute to growth in engagement interactions was raised..

The results of the third study made clear that the emotional support provided by teachers before Village@School (measured in a standardised situation) was positively related to the emotional support[74] provided by the teachers during the project, when controlling for teachers' scores on the other CLASS domains and for their attitudes (both assessed in the pre-measurement). However, teachers who scored higher on classroom organisation or who had a better attitude towards inquiry learning before the start of the project gave less emotional support during the implementation of Village@School (also when controlling for their pre-measurement scores on the other CLASS domains and their attitudes). This may

[74] For the emotional support during Village@School, school level but not class level variance was found.

indicate that teachers who establish a well-organised class life before the project become overwhelmed by a more chaotic environment in which pupils have more control over their learning environment. Teachers may become frustrated and experience difficulties when implementing the project-based S&T learning environment. This would be a normal reaction as "the scariest part of adopting an active, inquiry-based pedagogy for many teachers is the potential loss of control of individual students and control of the classroom" (Morgan & Slough, 2013, p. 100). Remarkably, however, teachers who are initially more enthusiastic about inquiry learning, about giving pupils opportunities for experimentation and initiative, may get easily disappointed, which can result in frustration and giving pupils less emotional support.

As the second study has shown, emotional support plays a role in pupils' growth in engagement. On the one hand, the average sensitivity of a teacher as measured four times during the research (pre, in-between and post) – when controlling for the other emotional support dimensions and for classroom organisation – leads to a higher growth in engagement in pupils. Pupils who are more comfortable asking the teacher questions, and who have teachers who are more sensitive and responsive to their needs and are more inclined to address their questions, issues and concerns (in comparison with other teachers), have a higher growth in engagement in comparison with other pupils. This is not that surprising, as teachers stated that some pupils got frustrated, for example because they could not find the necessary materials to build their construction. On the other hand, a negative connection was found between the average positive climate in the classroom – when controlling for the other emotional support dimensions – and pupils' growth in engagement. This is a surprising result, and different explanations were put forward when discussing the results of the study. One result indicated that in classes which are already characterised by a high positive climate, pupils may already receive more opportunities to work collaboratively, which may have resulted in a ceiling effect.

While the importance of sensitivity for pupils' (growth in) engagement in S&T learning environments has not yet been stressed in the literature, studies in other fields of education found positive effects of similar constructs (Birch & Ladd, 1997; Pianta, Nimetz, & Bennett, 1997). In the SDT (Deci, Vallerand, Pelletier, & Ryan, 1991; Ryan & Powelson, 1991) and self-system theory (Connell & Wellborn, 1991, in Roorda et al., 2011) three teacher supporting behaviours are discerned: showing involvement (i.e. caring for and expressing interest in the student), providing structure (i.e. setting clear rules and delivering on consequences), and supporting autonomy (i.e. giving students freedom to make their own choices and showing connections between schoolwork and students' interests). Of these three, teacher involvement seems to be the most important predictor of engagement (see Skinner & Belmont, 1993; Tucker et al., 2002). However, studies have also indicated the positive effects of a positive class climate on pupils' engagement, in combination with teacher sensitivity (e.g. Buyse et al., 2009), a result that was not found in Study 3.

Alongside the mixed results concerning the relation of emotional support to pupils' growth in engagement, a negative relationship with content understanding was found when controlling for teacher sensitivity and most of the instructional support dimensions. The discussion in the second study attributes this to the finding that in classes with a higher stimulation of content understanding before the project, pupils demonstrated higher levels of engagement. As such, the potential to grow in engagement was rather limited for pupils of these classes.

Finally, the effectiveness of the implementation of the project was not only detected via an investigation of pupils' growth in engagement, but also through an exploration of the possible growth in the teacher's style and attitudes. With regard to the evolution in teacher style, Study 3 has shown that throughout and after the implementation of Village@School, teachers gave more autonomy to pupils and provided a higher quality of feedback in their classes. On the other hand, throughout and after the implementation of the project there was more negativity (i.e. irritation, anger, sarcasm) visible in the interactions. However, the average score on the negative climate dimension for all teachers was still low. The increase in negative climate was explained by the fact that the project is challenging for teachers as it requires patience before one can see results, not only in terms of the visibility of the constructions on the plate, but also in terms of what was learnt by the pupils. On several occasions, teachers commented they believed that their pupils were not learning. In Village@School, pupils have autonomy throughout the learning process, which makes it difficult for teachers to keep track of each pupil's activities. With regard to the attitudes of teachers towards S&T (teaching), no growth was detected. We proposed the explanation that attitudes are difficult to change in a short time period and training should be explicitly focused on the change of these attitudes (van Aalderen-Smeets & Walma van der Molen, 2015). Furthermore, in another study conducted at CEGO (Van Cleynenbreugel et al., 2011), an increase in the attitude of teachers was found after a similar intervention in which a visit to a science centre (by the teacher) was included. This visit, and more specifically teachers' involvement during the visit, showed to be important for their attitude development. Self-reports of what they had gained from the sessions made it clear that the visit offered inspirational materials for teachers, and they were able to experience the importance of rich environments in learning about S&T in co-construction. Perhaps, teachers at first need a tangible experience in what S&T (teaching) can be, before they are able to implement rich S&T learning environments in their own classrooms. Nevertheless, in Study 3 most of the teachers already had relatively positive attitudes at the start of the project, which may have made a significant growth in attitudes difficult to detect.

An important conclusion that can be made from both studies is that the results are rather limited with regard to interactions. In the discussion in the second study, we have already pointed to the theoretical relevance of Regard for Student Perspectives, Analysis and Inquiry and Quality of Feedback in open-ended learning environments. An explanation for why we did not find correlations of these dimensions with pupils' growth in engagement was also discussed. While Regard for Student Perspectives and Analysis and Inquiry are more systemic parts of the arrangement of Village@School and the standardised S&T-assignments more generally – probably causing little variation between teachers with regard to their related interactions – this does not apply to the provided Quality of Feedback. Other studies in the field have stressed the importance of the effects of feedback. Wetzels (2015) found a positive relation between teachers' pedagogical-didactic strategies, such as scaffolding and asking questions that stimulate pupils to reason, and the growth in pupils' reasoning. As engagement also involves a cognitive dimension, it would have been likely to find a positive connection with the CLASS dimensions which anticipate this cognitive level. However, as was shown in Study 4, Quality of Feedback deals more with the amount of feedback that 'pushes' for learning than that which is geared to the learner's level of understanding and interests. It is in the dimensions Content Understanding and Analysis and Inquiry that the cognitive levels of content and activities are incorporated. More importantly, Content Understanding measures whether the teacher makes cognitively challenging content meaningful for pupils (see the discussion of the results in Study 4). While a reversed relation was detected between Content Understanding and pupils' growth in engagement in Study 2, a connection between Content Understanding and pupils' engagement ($\beta = .08$, $p = .003$), though not reported in this dissertation, was found while conducting the analyses. In the experiential education framework, engagement requires pupils to operate at the very limits of their capabilities, fostered by a speaking and challenging offer of materials, content and high-quality interventions geared to the learner's level of understanding and interests. As this requirement is not incorporated into all CLASS dimensions (see section 1 of this discussion), some results may not have been found.

In conclusion, the main positive results of Studies 2 and 3 indicated that pupils grew in their engagement, while teachers grew with regard to some aspects of their interactions. The differences between classes in growth in engagement after the project, as compared to before the start of the project, can be explained through the sensitivity of the teacher. In the literature on open-ended learning environments, it is often presupposed that pupils will profit 'automatically' from the benefits of these learning environments, without specifying the conditions under which the results apply. The teacher may play an important role in making a project-based learning environment like Village@School 'work'. His or her sensitivity did not grow over the course of the project, but the results made it clear that the initial emotional support provided by the teachers – of which sensitivity is part – was a predictor for the emotional support he or she provided during the Village@School project.

Limitations

Although some study-specific drawbacks have already been discussed, the following section will highlight more general limitations related to choices made during the doctoral research and the associated challenges for future research.

The first limitation of the intervention research relates to the decision to work without a control group. Therefore, it cannot be claimed that the growth in engagement was attributable to the intervention. While the use of a control group was taken into consideration, the manpower to conduct the data collection for this doctoral research project was limited. To conduct and analyse the six observations in the 34 classes was already a major organisational and practical challenge. Moreover, a decisive factor in this choice was that our interest primarily focused on the correlational data within the 'experimental group', for the potential explanation of teacher factors in the implementation of the project in the group who participated in the intervention.

Secondly, the studies in this doctoral research project are unique as, in comparison to previous studies, observations were conducted of a large group of teachers who were involved for eight (in the first wave) and seven (in the second wave) months respectively. In previous research – which mainly consists of case studies – observations have mostly been conducted on only a few teachers (e.g. Liljeström et al., 2013; Roth, 1998b; Trumbull, Scarano, & Bonney, 2006). Despite the fact that the sample is large – in terms of the number of observations – for the empirical analyses the power of some studies of this dissertation in particular for Study 3, was still low. This may also explain why in the factor analysis for the attitude questionnaire the factors were not clearly found. Some items had lower negative loadings, but the internal consistencies for attitudes toward science, technology, inquiry and design learning were still found when keeping them. However, these items could have been removed in order to obtain more less ambiguous scales for each attitude, which may have resulted in more outcomes being predicted by attitudes. Moreover, participation was voluntary for the schools and teachers in the sample, which may have resulted in more similarities than differences in the quality of their interactions. This may have resulted in the fact that variances on the different levels for the CLASS domains Classroom Organisation and Instructional Support could not be found in Study 3. It is therefore recommended that future research uses a randomly selected teacher sample.

Thirdly, a conscious choice was made to use an observational instrument: the CLASS. The CLASS instrument has proven to be reliable and valid, and measures similar aspects of interactions which are outlined in in the experiential education framework. Although the CLASS framework and the experiential education framework have overlapping characteristics, the focus of the CLASS framework in terms of generating (well-being) and engagement, appeared to be less demanding at some points. In the CLASS, pupil engagement is also seen as part of the quality of interactions itself, as a separate dimension, instead of a consequence of the class interactions. In future research the predictive validity of the CLASS for pupils' engagement as measured with the LIS-P – the experiential education instrument that was also used in the second study to measure engagement – has to be determined. A further development of the ASOS might encourage an exploration of the relation between teacher style (as conceptualised in the experiential education framework) and pupil engagement, using the data of this doctoral research.

Fourthly, we chose to use a general attitude questionnaire to assess teachers' attitudes towards S&T (teaching). However, in the questionnaire the concepts of science and technology had different meanings and were not comprehensively explained. It may be that some teachers had 'biology' in mind while others may have been thinking about 'physics' when filling in the science part of the questionnaire. A similar limitation of attitude instruments has also been reported in the literature (Palmer, 2004; Yates & Goodrum, 1990). More outcomes may have been found if a questionnaire specifically designed for the kinds of 'science' and 'technology' that are likely to occur in the standardised S&T assignments and in the Village@School project (e.g. biology to a lesser extent), was used. It is recommended that future research matches the content in the attitude instrument to the S&T learning environments with which the attitudes will be related.

Fifthly, a more practical problem with regard to the attitude questionnaire refers to the moment at which the questionnaires were filled in. In the post-measurement, some teachers completed the attitude questionnaire after the observation while most others completed it beforehand. In the pre-measurement, some teachers filled in the attitude questionnaire before instead of after the observations. Although it would have been better to have all teachers completing the attitude questionnaire at the same time, it was supposed attitude is a relatively stable construct that cannot be easily changed after doing one S&T- activity.

Sixthly, although the manpower to conduct the observations for this doctoral dissertation was limited, efforts were made to conduct observations which were as reliable as possible. With knowledge of the fact that it would have been better to ensure inter- instead of intra-rater reliability for the CLASS and the LIS-P over the whole intervention research, the raters – the PhD researcher herself in case of the CLASS and another researcher from the CEGO team in case of the ASOS – were thoroughly trained in conducting the observations. To score reliably with the CLASS, training was followed in Norway and a test was undertaken twice to become certified (one immediately after the training, and another one year after the first certification). In the pre-measurement of the first wave of data collection, the LIS-P *inter*-rater reliability could be determined (with a master thesis student as the other rater). With regard to the CLASS, it would have been better to score four cycles during each observation, as advised in a general guideline of the CLASS manual (Pianta et al., 2012). However, the manual also adds that the choice of the amount of cycles is dependent on the intentions of each research study. In this research it was not possible to score four cycles during each observation session because two classes had to be scored in one morning or afternoon to keep the planning realistic. In Study 2 it was possible to overcome this limitation by using an average score of the CLASS- interactions for all four observations.

A seventh limitation is that a mix of live- and video- scored observations were used for the empirical studies in the second and third study. While the initial aim was to score all observations in person, it was not always possible to do so. Although teachers were asked to take 45 minutes to conduct each observation session, some teachers stopped earlier or organised a clean-up session in the last 15 minutes. In that case, it was not possible to score a second cycle live with the CLASS. Exceptionally, doubt existed about the score to give on one or more dimensions, which made us come to the decision to score that cycle or dimension again using the video recording. As observations were initially scored with the CLASS in the pre-measurement of the first wave, it was decided to rescore these observations using the video. The Negative Climate dimension was rescored on the video for all observations in the first wave. Initially, this dimension seemed to be irrelevant because of the apparently low variation between teachers, but it was scored later on to be able to calculate a composite score for classroom organisation. Due to difficulties surrounding the time it took to travel to schools, it was decided that observations for the pre- and post-measurement of the second wave should be scored using the video[75]. While a t-test was conducted to test for the differences between live and video observations, this was only done for the two observations during Village@School (domain scores), but could not be done for all four observation occasions, as all final scores for the observations in the pre-measurement were video-scores and not a mix of live and video scores. However, the dimension scores of the observation occasions were used in the data analyses of Study 2. In a few cases technical problems with the camera occurred (e.g. the teacher's microphone did not work). While the camera was supervised by the CLASS rater, it was not possible to constantly check the camera while observing.

[75] The researcher who scored pupils' engagement made the video-recordings.

An eighth limitation was that, to exclude as much subjectivity as possible in the scoring of the observations, the individual coaching sessions were conducted by a colleague. Although the raters were involved in the organisation of the conference at the start and the workshops throughout, and the CLASS rater also conducted the teacher interviews, efforts were made to keep the coaching role separate from the assessment role. This was sometimes a challenge, as teachers showed to be insecure during the implementation of the project and sought affirmation, even before or after observations in which no coaching session was planned. An important reflection in this respect is that combining intensive data collections with coaching in action research of a relatively large – compared with other studies – group of teachers presents a challenge. In particular, this may be the case when implementing an open-ended (S&T) learning environment like Village@School. Only when there are opportunities to involve different data collectors and coaches will it be possible to provide more support for teachers. Also related to this support, the question remains of how to set up coaching sessions for teachers in the field of S&T education, as the literature is not yet clear when it comes to the role of the teacher. One can only coach in the direction of generally hypothesised working elements in these S&T learning environments. Furthermore, from the teachers' viewpoint, researchers are often seen as 'experts' in the field. Although we aimed to conduct a dialogue between researchers and practitioners to improve class practice, in future research more 'equal' conversations could be more explicitly embedded in the intervention. In that vein, the organisation of the individual coaching sessions could be improved; for example by making use of video feedback coaching, a technique elaborately investigated and implemented to stimulate interactions in S&T in primary school by Wetzels (2015). The participating teachers in Wetzels' research also reported greater confidence and enjoyment in incorporating science teaching into their classrooms. In the present study observations were recorded with a camera with a wireless microphone. The video recorded observations were used for the sake of research, but they could also be used as a means to coach teachers throughout the trajectory. It is noteworthy that the individual coaching sessions conducted in the research were also recorded, which makes it possible to determine the stronger and weaker points in these coaching sessions in future research. While most of the coaching sessions were done with each teacher individually, in some cases the coaching sessions were conducted with two teachers from the same school participating in the project. In future research, it would be interesting to explore whether teachers experience the participation of a colleague teacher from the same school as supportive when implementing a challenging learning environment like Village@School. This could be done by using a Lesson Study model in which two or more teachers purposely work together from the beginning to plan, organise and share their experiences with regard to the project (Kotelawala, 2012). Previous research has shown that communities of inquiry within the teacher's own classroom or their colleagues' classrooms are more promising than professional development without opportunities for collaboration (Ball & Cohen, 1999, in Kotelewala, 2012; Ingvarson, Meiers, & Beavis, 2005). The practice of Lesson Study offers such a community of inquiry (Kotelewala, 2012).

A ninth limitation is that the intervention as well as the data collection of this research was highly demanding for teachers. They were asked to implement a new, unknown innovative learning environment in the fields of S&T, which are fields they were often not familiar with. The different data collection methods – observations, the administration of a teacher and student questionnaire and test, interviews and teacher diaries[76] – and the participation in the conference at the start and in the workshops required an effort from teachers. A few teachers experienced the video-recorded observations as threatening. To put these teachers at ease and to guarantee the ecological validity of the studies conducted, it was regularly emphasised that the observations were recorded for the research and that is was not aimed to judge individual teachers on their practice.

Finally, it is important to note that this doctoral research was ambitious from the start. Initially, and next to focusing on the interactions at the micro-level, it also aimed to fully investigate the broader conditions of the Village@School project (e.g. general amount of initiative provided and richness of materials). Therefore, additional data were collected via teacher interviews and teacher diaries. However, time restrictions did not allow the researchers analyse these data. In the future, teacher diaries and interviews could be used to gain a substantial insight into the specific processes and the achievement of the project aims.

[76] The test (Minitest - 'Eye for Science and Technology'), teacher interviews and diaries, and a pupil questionnaire, were not used in this dissertation.

Directions for future research

Firstly, this research made it clear that the implementation of open-ended S&T learning environments is not self-evident. With this research a call is made to further investigate aspects of class interactions, and especially the often overlooked affective elements of these interactions. Particularly, the role of the sensitivity of the teacher for pupils' growth in engagement in open-ended S&T learning environments should be further investigated. We need to examine the way in which this sensitivity plays a role. Via teacher and pupil questionnaires, it may be possible to determine how pupils experience the awareness and responsiveness of the teacher as a help when making their way in these (generally more chaotic) S&T learning environments. It is also important to determine whether sensitivity is more important for some pupils in comparison to others. For example, it is necessary to identify whether a sensitive teacher works protectively for pupils showing off-task behaviour in open-ended S&T learning environments (Soares & Vannest, 2013). A teacher can help to identify parts of the project that the pupil is responsible for and able to complete.

Secondly, we recommend further examination of the positive relation between the role of stimulating content understanding in interactions and pupils' engagement, as this was not discussed in this dissertation. In the CLASS the concept of content understanding is covered very broadly. It would be interesting to investigate whether certain aspects are relatively more important in comparison to others. Classes could score high on content understanding for different reasons: it could be due to the reached depth regarding the content of the dialogues, or it could be a result of using real-world materials. Future research should be encouraged to find out which of these aspects have more impact on pupil engagement. In other studies concerning science education, researchers went a step further by unravelling interactions by making use of discourse analysis (Mercer, 2010). Researchers at CEGO have had the opportunity to do this. In the past, a content analysis of the interventions scored with an earlier version of the ASOS was made (Vervoort, 2011). In line with a further development of the instrument, it would be interesting to thoroughly analyse interventions that fit into the three dimensions of the teacher style instrument, and relate these to pupils' engagement.

Thirdly, not only should future research investigate the interventions of the teacher, but the interactions among pupils also deserve special attention. During the observations, in some classes in which pupils had less positive relations with teachers and/or received little stimulation from the teacher, pupils still appeared motivated to build the village. Therefore it would be interesting to determine whether the Village@School project in itself and/or the interactions among pupils may work to stimulate high levels of engagement, despite teacher interventions. By using a control group, and an instrument that exclusively focuses on the interactions among pupils, these hypotheses could be investigated. The research on self-organising systems of learning (SOLEs) is inspirational in this context (Mitra & Dangwal, 2010).

Finally, the hypothesis that teachers with an initially higher attitude towards inquiry learning and a higher level of classroom organisation become easily disappointed or frustrated when they have to cope with the difficulties of implementing an open-ended S&T learning environment should be further investigated. It would be interesting to investigate which factors cause this frustration or disappointment. Moreover, as discussed in Study 3, Van Aalderen-Smeets et al. (2011) proposed a new theoretical framework for attitudes in which they distinguish an 'Anxiety' dimension in the affective component of attitudes, and a new 'Perceived Control' component, subdivided into the dimensions 'Self-efficacy' and 'Context Dependency'. In the last two dimensions, teachers' beliefs about and feelings of being in control of executing particular behaviours (self-efficacy) as well as their beliefs and feelings about external (contextual) factors that make them feel in control, are evaluated. Based on this framework, Van Aalderen-Smeets and Walma van der Molen (2013) developed and validated a new attitude instrument, called the Dimensions of Attitude towards Science instrument. The new theoretical framework and instrument provide an interesting perspective from which to dig deeper into teacher's attitudes towards S&T (and its teaching) in future research. In addition, it is recommended that researchers explore more aspects of the teacher's competence profile; also SMK (Abell, 2007; Gess-Newsome, 1999; both in Friedrichsen et al., 2010) and PCK (Abell, 2007, in Friedrichsen et al., 2010; Davis et al., 2006) may also play a role in the way teachers interact with pupils in open-ended S&T learning environments.

Implications for educational practice

The main result – that the sensitivity of the teacher plays an important role for pupils' growth in engagement – is good news for educational practice. Although a lot of teachers are insecure when they start teaching S&T, and despite the fact that further research is needed, the results show that teachers who do not have experience or expertise in teaching S&T are capable of increasing pupils' engagement in open-ended S&T learning environments. At first, it is important to be aware of and responsive to learners' needs, problems and struggles while they are exploring and designing in the open-ended S&T learning environment. While this message seems to be simplistic, in previous research similar conclusions were made. Mitra (2014, p. 551) for example shows that only introducing "an affectionate and admiring, but not knowledgeable, adult" results in increasing levels of learning with regard to the basic concepts of biotechnology in SOLEs of primary school pupils. He points to the positive effects of "the 'grandmother's method': stand behind, admire, act fascinated and praise" (Mitra, 2014, p. 551). However, we have to be aware that various aspects in this dissertation were measured with CLASS dimensions – which also contained other aspects – that were not, or were negatively, connected to pupils' growth in engagement. The aspects of 'praising' and 'admiring' in this PhD research were measured with the Positive Climate dimension, which was negatively related to pupils' growth in engagement, and with the Quality of Feedback dimension, which was not related to this growth. The fascination or interest of the teacher during the activities was measured in the Instructional Learning Formats dimension, on which no analyses have been conducted. In comparison to the grandmother's method, the Teacher Sensitivity dimension – which showed to be positively related to pupils' growth in engagement – may indicate a more active and genuine emotional involvement of the teacher in what pupils are doing (with indicators like awareness, responsiveness for students' needs, effectiveness in addressing problems and student comfort). Future research in educational practice should give a profound insight into how active the emotional involvement of the teacher has to be in order to have the desired effects in open-ended S&T learning environments.

However, it seems difficult for some teachers to provide this emotional support in all circumstances. Teachers who are used to smooth classroom organisation may perceive a learning environment like Village@School to be chaotic, not only because they can no longer predetermine what will be learnt and how this will be done, but also because of the fundamentally new type of classroom organisation (Morgan & Slough, 2013). This is not surprising, as (1) teachers are not always educated (enough) to set up such learning environments or able to coach pupils in these learning environments, and (2) Flemish and Dutch teachers experience a lot of pressure from the school inspectorate to reach the National educational standards. It is often easier to use clear-cut methods in which the goals are set beforehand instead of using an open-ended learning environment in which the curriculum ‘emerges’ (Laevers, 2011), and in which pupils themselves choose what they are going to learn and how they will do this. This may give teachers the perception of not being in control of the situation. With regard to organisation, a shift of focus is also needed when implementing open-ended S&T learning environments. Instead of rules and procedures which emphasise listening to the teacher and following predetermined steps to solve problems, rules and procedures which emphasise student engagement, decision making, and problem solving need to be designed and implemented (Morgan & Slough, 2013). Examples of these rules for students are: “Record all design ideas in your lab notebook/journal” and “All students are responsible for all phases of the project, regardless of their temporary roles within the group” (Morgan & Slough, 2013, p. 100).

Therefore, it may be necessary to invest in pre-service and in-service teacher education to make (future) teachers aware of the opportunities that exist to design or use learning environments which have the potential to increase pupil engagement. When doing so, it is important that teachers learn to observe how and what pupils are learning, so that they can explore the possibilities for increased pupil engagement and its benefits over traditional instruction. In this respect, the further development of user-friendly instruments which can be used by both researchers and teachers – such as the ASOS – may be fruitful. Moreover, it is important to note that pre-packaged ‘one size fits all’ instructional kits cannot be provided, because this would conflict with the basic principles of open-ended S&T learning environments. Perhaps it is important to focus on the general capacity of teachers to act in a variety of learning environments. It is important that they get used to such open ended S&T learning environments, as these represent the most accepted ways to learn S&T in primary school. This dissertation demonstrated that teachers have begun to grow in this capacity. They have learnt to trust and give autonomy and responsibility to their pupils – as demonstrated by teachers’ growth after the intervention – but a next step may still have to be taken. When acting in such open-ended S&T learning environments, they have to believe that the pupils will be able to get to interact with and engage in the learning environment without step-by-step procedures (Morgan & Slough, 2013). Professional development in S&T deserves special attention as individual teachers are critically important for science education and reform efforts (Borko, 2004; Cobern & Loving, 2002; Desimone, 2009; Haney & Lumpe, 1995; Nye, Konstantopoulos, & Hedges, 2004). As well as professional development, incorporating S&T more often into the classroom activities may improve the teacher’s role in using organisational forms and interventions which make learners co-owners of their development (Laevers, 2011). According to Laevers (2011), S&T education ‘automatically’ asks for reality to be brought in, in the form of genuine, concrete and complex materials. This complex physical reality will generate new questions from pupils, which have to be explored and investigated, as the teacher will not know all of the answers. As such, the teacher is ‘forced’ to support learners’ exploration of and design in this reality.

Conclusion

This dissertation is a contribution not only to S&T education, but also to the broader educational research field. The extensive literature review meets the needs of researchers to find suitable instruments, scales, items and questions for their own studies, and at the same time calls on them to develop new instruments and to invest in the description of their operationalisation. Taken together, the studies in this dissertation will hopefully inspire future researchers to conduct more research in the real class context; not only to further investigate and develop the role of the teacher, in all its facets, but also to study the engagement of pupils in challenging, open-ended S&T learning environments which are also new for them. When doing so, teachers and their colleagues have to be genuinely involved, not only to observe opportunities for improvement but also to discover the power of the learning environment and their own role therein. It is clear that such research will be a challenge – from choosing the right instruments, to conducting intensive data collection in large samples of teachers – but this should not hold back future researchers as they contribute to improving the quality of (S&T) education.

References

Abd-El-Khalick, F. , & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. *International Journal of Science Education, 22,* 665-701.

Aiken, L., & West, S. (1991). Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage.

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology, 52,* 27-58.

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology, 32,* 665-683.

Ajzen, I. , & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. Englewood-Cliffs, NJ: Prentice-Hall.

Alake-Tuenter, E., Biemans, H. J. A., Tobi, H., & Mulder, M. (2013). Inquiry-based science teaching competence of primary school teachers: A Delphi study. *Teacher and Teacher Education, 35,* 13-14.

Alake-Tuenter, E., Biemans, H. J. A., Tobi, H., Wals, A. E. J., Oosterheert, L., & Mulder, M. (2012). Inquiry-based science education competencies of primary school teachers: a literature study and critical review of the American National Science Education Standards. *International Journal of Science Education, 34,* 2609-2640.

Alexander, R. (2005). *Towards dialogic teaching* (Vol. 2). UK: Dialogos.

Alexander, R. (2008). Towards dialogic teaching: Rethinking classroom talk. 4th ed. York, England: Dialogos.

Alexander, P. A., & Wade, S. E. (2000). Contexts that promote interest, self-determination, and learning: Lasting impressions and lingering questions. *Computers in Human Behavior, 16,* 349-358.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? Journal of Educational Psychology, 103, 1-18.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333* , 1034-1037.

Allen, J. P., Gregory, A., Mikami, A. Y., Lun, J., Hamre, B., & Pianta, R. C. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the CLASS-S. *School Psychology Review, 42,* 76-98.

Allen, J. P., Hauser, S. T., Bell, K. L., & O' Connor, T. G. (1994). Longitudinal assessment of autonomy and relatedness in adolescent-family interactions as predictors of adolescent ego development and self-esteem. *Child Development, 65,* 179-194.

Allen, J., Kuperminc, G., Philliber, S., & Herre, K. (1994). Programmatic prevention of adolescent problem behaviors: The role of autonomy, relatedness, and volunteer service in the teen outreach program. *American Journal of Community Psychology, 22,* 617-638.

Allen, J. P. , Marsh, P., Mcfarland, C., Mcelhaney, K. B., Land, D. J., Jodl, K. M., Peck, S. (2002). Attachment and autonomy as predictors of the development of social skills and delinquency during midadolescence. *Journal of Consulting and Clinical Psychology, 70,* 56-66.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333,* 1034-1037.

Amaral, O. M., Garrison, L., & Klentschy, M. (2002). Helping English learners increase achievement through inquiry-based science instruction. *Bilingual Research Journal, 26,* 213-239.

American Association for the Advancement of Science (2002). *Criteria for evaluating the quality of Instructional support.* Retrieved from http://www.project2061.org/publications/textbook/mgsci/summary/criteria.htm

Anastasi, A., & Urbina, S. (1996). *Psychological testing* (7th ed.). New York: Macmillan.

Andersen, A. M., Dragsted, S., Evans, R. H., & Sorensen, H. (2004). The relationship between changes in teachers' self-efficacy beliefs and the science teaching environment of Danish first-year elementary teachers. *Journal of Science Teacher Education, 15,* 25-38.

Appleton, K., & Kindt, I. (1999). Why teach primary science? Influences on beginning teachers'practices. *International Journal of Science Education, 21,* 155-168.

Asma, L. J. F., Walma van der Molen, J. H., & Van Aalderen-Smeets, S. I. (2011). Primary teachers'attitudes towards science: Results of a focus group study. In M.J. de Vries, H. van Keulen, S. Peters, & J.H. Walma van der Molen (Eds.), *Professional development for primary teachers in science. The Dutch VTB-Pro project in an international perspective* (Chapter 8, pp. 89-106). Rotterdam, The Netherlands:

Sense Publishers. Retrieved from http://link.springer.com.kuleuven.ezproxy.kuleuven.be/book/10.1007/978-94-6091-713-4/page/1

Atman, C. J., Kilgore, D., & McKenna, A. (2008). Characterizing design learning: A mixed-methods study of engineering designers' use of language. *Journal of Engineering Education, 97,* 309-326.

Avraamidou, L., & Zembal-Saul, C. (2005). Giving priority to evidence in science teaching: A first-year elementary teacher's specialized knowledge and practice. *Journal of Research in Science Teaching, 42,* 965-986.

Baker, W. P., Lang, M. , & Lawson, A. E. (2002). Classroom management for successful student inquiry. *The Clearing House, 75,* 248-252.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: what makes its special? *Journal of Teacher Education, 59,* 389-407.

*Banks, D. L., Elser, M., & Saltz, C. (2005). Analysis of the K-12 component of the Central Arizona-Phoenix Long-Term Ecological Research (CAP LTER) project 1998 to 2002. *Environmental Education Research, 11,* 649-663.

Barak, M. (2004). Issues involved in attempting to develop independent learning in pupils working on technological projects. *Research in Science & Technological Education, 22,* 171-183.

Barak, M., & Doppelt, Y. (1999). Integrating the CoRT program for creative thinking into a project-based technology curriculum. *Research in Science and Technological Education, 17,* 139-151.

Barak, M., & Doppelt, Y. (2000). Using portfolios to enhance creative thinking. *Journal of Technology Studies, 26,* 16-24.

Barak, M., Eisenberg, E., & Harel, O. (1995). What's in the calculator? An introductory project for technology studies. *Research in Science & Technological Education,* 12, 147-154.

Barak, M., & Shachar, A. (2008). Projects in technology education and fostering learning: The potential and its realization. *Journal of Science Education and Technology, 17,* 285-296.

Barak, M., & Raz, E. (2000). Hot air balloons: Project-centered study as a bridge between science and technology education. *Science Education, 84,* 27-42.

Barmby, P., Kind, P. M., & Jones, K. (2008). Examining changing attitudes in secondary school science. *International Journal of Science Education,* 30, 1075-1093.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., & Bransford, J. D. (1998). Doing with understanding: Lessons from research on problem and project-based learning. *Journal of the Learning Sciences, 7*, 271-311.

Barros, S. S., & Elia, M. F. (1998). Physics teacher's attitudes: How do they affect the reality of the classroom and models for change? In A. Tiberghien, E. L. Jossem, & J. Barojas (Eds.), Connecting research in physics education with teacher education (section D2) (pp. 267-278). Ohio: ICPE.

Barth, R. S. (1972). *Open education and the American school.* New York, NY: Shoken
Books.

*Baxter, B. K., Jenkins, C. C., Southerland, S. A., & Wilson, P. (2004). Using a multilevel assessment scheme in reforming science methods courses. *Journal of Science Teacher Education, 15*, 211-232.

Bell, B., & Gilbert, J. (1994). Teacher development as professional, personal, and social development. *Teacher and Teacher Education, 10*, 483-497.

Bennett, J., Rollnick, M., Green, G., & White, M. (2001). The development and use of an instrument to assess students' attitude to the study of chemistry. *International Journal of Science Education, 23*, 833-845.

Benoit, C. (2015). *Leerkrachtstijl in de context van het lager onderwijs. Optimalisering van de Adult Style Observation Schedule (ASOS)* [Teacher style in the context of elementary education. Optimization of the Adult Style Observation Schedule (ASOS)].

Unpublished Master's thesis, University of Leuven, Faculty of Psychology and Educational Sciences, Belgium.

Bhattacharyya, S., Volk, T., & Lumpe, A. (2009). The influence of an extensive inquiry-based field experience on pre-service elementary student teachers' science teaching beliefs. *Journal of Science Teacher Education, 20*, 199-218.

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*, 61-79.

Bitner, L. B. (1994). *Revised science attitude scale for preservice elementary teachers: Re-examined.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Anaheim, CA.

Blair, C. (2003). School readiness as propensity for engagement: Integrating cognition and emotion in a neurobiological conceptualization of child functioning at school entry. *American Psychologist, 57* , 111-127.

Blank, R. K., de las Alas, N., & Smith, C. (2008). *Does teacher professional development have effects on teaching and learning?* Evaluation findings from programs in 14 states. Washington, DC: Council of Chief State School Officers.

Blatchford, P. (1992). Children's attitudes to work at 11 years. *Educational Studies, 18*, 107-118.

*Bodzin, A. M., & Beerer, K. M. (2003). Promoting inquiry-based science instruction: The validation of the Science Teacher Inquiry Rubric (STIR). Journal of *Elementary Science Education, 15,* 39-49.

Bodzin, A., & Cates, W. M. (2002). *Web-based inquiry for learning science (WBI) instrument manual.* Retrieved from www.lehigh.edu/~amb4/wbi/wbi-v1_0.pdf.

Boeckx, K., & Boonen, I. (2009). *Klasklimaat. Onderzoek naar de betrouwbaarheid en validiteit van de Checklist Klasklimaat* [Class climate. Investigation of the reliability and validity of the Checklist Class climate], Unpublished dissertation, University of Leuven, Faculty of Psychology and Educational Sciences, Belgium.

Boekaerts, M. (2002). Bringing about change in the classroom: strengths and weaknesses of the self-regulated learning approach – EARLI Presidential Address, 2001. *Learning and Instruction, 12,* 589-604.

Barron, B., Schwartz, D., Vye, N., Moore, A., Petrosino, L., & Bransford, J. (1998). Doing with understanding: Lessons from research on problem and project based learning. *Journal of the Learning Sciences, 7*(3), 271-311.

Blumenfeld, P., Soloway, E., Marx, R., Krajcik, J., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26,* 369-398.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15.

Brandt, R. (1998). *Powerful learning.* Alexandria, VA: Association for Supervision and Curriculum Development (ASCD). Retrieved from https://books.google.be/books?hl=nl&lr=&id=bD08AcMW7RIC&oi=fnd&pg=PT5&ots=D1ejE7QeGj&s

Bransford, J. D., Brown, A. L., & Cocking, R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

*Braund, M., & Leigh, J. (2013). Frequency and efficacy of talk-related tasks in primary science. *Research in Science Education, 43,* 457-478.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology, 47,* 1191-1205.

Brouwer, N., & Korthagen, F. (2005). Can teacher education make a difference? *American Educational Research Journal, 42,* 153-224.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Buaraphan, K. (2012). Embedding nature of science in teaching about astronomy and space. *Journal of Science Education and Technology, 21,* 353-369.

Buyse, E., Snoeck, G., Laevers, F., Bertrands, E., Declercq, B., Van Gorp, K., Verheyden, L., Blaton, K. (2009). *Fundamenteel diepte-onderzoek naar krachtige GOK-leeromgevingen. Onderzoeksluik betrokkenheid & welbevinden.* [Fundamental in-depth research to powerful GOK-learning environments. Research section with regard to involvement and well-being].

Carleton, L. E., Fitch, J. C., & Krockover, G. H. (2008). An inservice teacher education program's effect on teacher efficacy and attitudes. *The Educational Forum, 72*(1), 46-62.

Carlsen, W. S. (1993). Teacher knowledge and discourse control: quantitative evidence from novice biology teachers' classrooms. *Journal of Research in Science Teaching, 30,* 417-481.

Catts, H., Fey, M., Zhang, X., & Tomblin, J. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech & Hearing Services in Schools, 32,* 38.

Champagne, A. B., Gunstone, R. F., & Klopfer, L. E. (1985). Instructional consequences of students' knowledge about physical phenomenon. In L. H. T. West, & A. L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 61-68). New York: Academic Press.

ChanLin, L.-J. (2008). Technology integration applied to project-based learning in science. *Innovations in Education and Teaching International, 45*(1), 55-65.

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching, 44*(6), 815-843.

Choi, S., & Ramsey, J. (2010). Constructing elementary teachers' beliefs, attitudes, and practical knowledge through an inquiry-based elementary science course. *School Science and Mathematics, 109,* 313-324.

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing inter-rater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*(2), 127-137.

Clark, L. (1999). *Learning from the field: The journey from post-positivist to constructivist methods.* Paper presented at the meeting of the International Communication Association, San Francisco.

Coates, D., & Wilson, H. (2003). *Challenges in primary science: Meeting the needs of able young scientists at Key Stage 2.* London: NACE/Fulton.

Cobern, W. W., & Loving, C. C. (2002). Investigation of preservice elementary teachers' thinking about science. *Journal of Research in Science Teaching, 39*(10), 1016-1031.

Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research, 77*, 113-143.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cornell, N., & Clarke, J. (1999). The cost of quality: evaluating a standards-based design project. *National Association for Secondary School Principals Bulletin, 83*(603), 91-99.

Coulson, R. (1992). Development of an instrument for measuring attitudes of early childhood educators towards science. *Research in Science Education, 22*, 101-105.

Crawford, B. A. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching, 37,* 916-937.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart, & Winston.

Csikszentmihalyi, M. (1988).The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. Csikszentmihalyi (Eds.), *Optimal Experience* (pp. 15-35). Cambridge, UK: Cambridge University Press.

Cross, P (1998). Why learning communities? Why now? *About Campus, 3,* 4-11.

Csikszentmihalyi, M. (1999). If we are so rich, why aren't we happy? *American Psychologist, 54,* 821-827.

Curtis, D. (2002). The power of projects. *Educational Leadership, 60*(1), 50-53.

Damhuis, R., & De Blauw, A. (2011). High quality interaction in science and technology education. In: de Vries, M. J., van Keulen, H., Peters, S., & van der Molen, J. W. (Eds.), *Professional development for primary teachers in science and technology. The Dutch VTB-Pro project in an international perspective* (pp. 199-216). Rotterdam: Sense Publishers. Retrieved from http://link.springer.com.kuleuven.ezproxy.kuleuven.be/book/10.1007/978-94-6091-713-4/page/1

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). State of the profession: study measures status of professional development. *Journal of Staff Development, 30,* 42-50.

Davidson, J.E., & Sternberg, R.J., (2003). *The psychology of problem solving.* Cambridge: Cambridge University Press.

Davies, T. (2000). Confidence! Its role in the creativity teaching and learning of design and technology. *Journal of Technology Education, 12,* 18-31.

Davis, E. A.,Petish, D., & Smithey, J. (2006). Challenges new science teachers face. *Review of Educational Research, 76,* 607-651.

Daymon, C., & Holloway, I. (2011). Qualitative research methods in public relations and marketing communications (2nd ed.). London: Routledge.

Deci, E.M., & Ryan, R.M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behaviour. *Psychological Inquiry, 11,* 227-268.

Deci, E.L., Schwartz, A., Sheinman, L., & Ryan, R.M. (1981). An instrument to assess adults' orientations toward control versus autonomy in children: Reflections on intrinsic motivation and perceived competence. *Journal of Educational Psychology, 73,* 642-650.

Deci, E. L., Vallerand, R.J., Pelletier, L.G., & Ryan, R.M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist, 26,* 325-346.

De Fraine, B. (2004). Onderwijseffectiviteit: overzicht van de onderzoeksliteratuur binnen een evoluerend domain. [Educational effectiveness: overview of the research literature in an evolving domain] In J. Van Damme, G. Van Landeghem, B. De Fraine, M.-C.

Opdenakker, & P. Onghena (Ed.), *Maakt de school het verschil? [Makes the school the difference?]* Leuven, Belgium: Acco.

De Jong, O. (2003). Exploring science teachers' pedagogical content knowledge. In D. Psillos, P. Kariotoglou, V. Tselfes, E. Hatzikraniotis, G. Fassoulopoulos, & M. Kallery (Eds.), *Science Education research in the knowledge-based society.* Dordrecht: Kluwer Academic Publishers.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, *34*, 12-25.

Department of Education and Skills (2012). *Science in the primary school 2008: inspectorate evaluation studies.* Marlborough Street: Evaluation and Support and Research Unit Inspectorate, DES.

de Kruif, R. E. L., McWilliam, R. A., Ridley, S. M., Wakely, M. B. (2000). Classification of teachers' interaction behaviors in early childhood classrooms. *Early Research Quarterly*, *15*(2), 247-268.

Desimone, L.M. (2009). Improving impact studies of teachers' professional development: toward better conceptualizations and measures. *Educational Researcher*, *38*, 181-199.

Destoforges, S., & Cockburn, A. (1987). *Understanding the mathematics teacher: a study of practice in first schools.* London: Falmer Press.

du Chatenier, E., Verstegen, J. A. A. M., Biemans, H. J. A., Mulder, M., & Omta, O. S. W. F. (2010). Identification of competencies for professionals in open innovation teams. *R&D Management*, *40*, 271-280.

De Winter, V., Stas, T., & Thys, M. (2013). Dorp Op School: Haal de sterren in je klas naar boven! Informatiebundel [Village@School: Bring the stars of your class to the surface! Information guide]. Leuven: CEGO.

De Winter, V., Van Cleynenbreugel, C., Buyse, E., & Laevers, F. (2010). *Leerkrachtprofielen en onderwijs in wetenschap en techniek in het basisonderwijs: Werkzame bestanddelen voor deskundigheidsbevordering en attitudeverandering. VTB-Pro aanvullend onderzoek 2009-2010. Eindverslag. [Teacher profiles and education in S&T in elementary school. Working constituents for expertise advancement and change. Final Report.]* Leuven: CEGO.

Diehl, W., Grobe, T., Lopez, H., & Cabral, C. (1999). *Project-based learning: A strategy for teaching and learning.* Boston, MA: Center for Youth Development and Education, Corporation for Business, Work, and Learning.

Dietz, C. M., & Davis, E.A. (2009). Pre-service elementary teachers' reflection on narrative images of inquiry. *Journal of Science Teacher Education, 20,* 219-243.

Donaldson, C. (2007). Group work in elementary science: Towards organisational principles for supporting pupil learning. *Learning and Instruction, 17,* 549-563.

Donahue, T. P., Lewis, L. B., Price, L. F., & Schmidt, D. C. (1998). Bringing science to life through community-based watershed education. *Journal of Science Education and Technology, 7,* 15-23.

Doppelt, Y. (2003). Implementation and assessment of project-based learning in a flexible environment. *International Journal of Technology and Design Education, 13,* 255-272.

Doppelt, Y. (2005). Assessment of project-based learning in a Mechatronics context. *Journal of Technology Education, 16*, 7-24.

Doppelt, Y. (2009). Assessing creative thinking in design-based learning. *International Journal of Technology and Design Education, 19,* 55-65.

Doppelt, Y., & Barak, M. (2002). Pupils identify key aspects and outcomes of a technological learning environment, *Journal of Technology Studies, 28*, 12-18.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*, 5-12.

Doumen, S., Koomen, H. M. Y., Buyse, E., Wouters, S., Verschueren, K. (2012). Teacher and observer views on student-teacher relationships: Convergence across kindergarten and relations with student engagement. *Journal of School Psychology, 50,* 61-76.

Eagly, A., & Chaiken, S. (1993). *The psychology of attitudes.* Belmont, CA: Wadsworth group/Thomson Learning.

Edmunds, J., Ntoumanis, N., & Duda, J.L. (2008). Testing a self-determination theory-based teaching style intervention in the exercise domain. *European Journal of Social Psychology*, *38*, 375-388.

Eick, J. C. (2011). Use of the outdoor classroom and nature-study to support science and literacy learning: A narrative case study of a third-grade classroom. *Journal of Science Teacher Education, 23,* 789-803.

Engeström, Y. (1992). *Interactive expertise: Studies in distributed working intelligence.*

Erdogan, I., & Campbell, T. (2008). Teacher questioning and interaction patterns in classrooms facilitated with differing levels of constructivist teaching practices. *International Journal of Science Education, 30,* 1891-1914.

Eshach, H., Dor-Ziderman, Y., & Yefroimsky, Y. (2014). Question asking in the science classroom: Teacher attitudes and practices. *Journal of Science Education and Technology, 23,* 67-81.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38,* 47-65.

Fallik, O., Eylon, B.-S., & Rosenfeld, S. (2008). Motivating teachers to enact free-choice project-based learning in science and technology (PBLSAT): Effects of a professional development model. *Journal of Science Teacher Education, 19,* 565-591.

Flynn, E., Pine, K., & Lewis, C. (2006). The microgenetic method: Time for change? *The Psychologist, 19,* 152-155.

*Forbes, C. T., & Davis, E. A. (2008). The development of preservice elementary teachers' curricular role identity for science teaching. *Science Education, 92,* 909-940.

*Forbes, T. C., & Davis, A. E. (2011). Operationalizing identity in action: A comparative study of direct versus probabilistic measures of curricular role identity for inquiry-based science teaching. *International Journal of Science and Mathematics Education, 10,* 267-292.

Fortus, D., Dershimer, R. C., Krajcik, J., Marx, R. W., & Mamlok-Naaman, R. (2004). Design-based science and student learning. *Journal of Research in Science Teaching, 41,* 1081-1110.

Fraser, B. J. (1994). Research on classroom and school climate. In D. Gabel (ed.), *Handbook of research on science teaching and learning* (pp. 493-541). New York, NY: Macmillan.

Fraser, B.J. (1998). Classroom environment instruments: Development, validity and applications. *Learning Environments Research, 1,* 7-33.

Fraser, B. J. (2012). Classroom learning environments: Retrospect, context and prospect. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (Chapter 79, pp. 1191-1239). Dordrecht: Springer. Retrieved from http://link.springer.com/chapter/10.1007%2F978-1-4020-9041-7_79

Fraser, B. J., & O'Brien, P. (1985). Student and teacher perceptions of the environment of elementary school classrooms. *The Elementary School Journal, 85,* 567-580.

Fredricks, J.A., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74,* 59-109.

Friedrichsen, P., Van Driel, J.H., & Abell, S.K. (2010). Taking a closer look at science teaching orientations. *Science Education, 95,* 358-376.

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82, 300-329.

Gable, R. K., & Wolf, M. B. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings.* Boston: Kluwer Academic Publishers.

Galton, M. (1980). Ins*ide the primary classroom.* London: Routledge & Kehan Paul.

Galton, M., & MacBeath, J. with Page, C., & Steward, S. (2002). *A life in teaching? The impact of change on primary teachers' working lives.* Retrieved from http://www.educ.cam.ac.uk/download/NUTreport.pdf.

Germann, P.J., & Aram, R.J. (1996). Student performances on the science processes of recording data, analyzing data, drawing conclusions, and providing evidence. *Journal of Research in Science Teaching, 33*, 773-98.

Gess-Newsome, J. (1999). Secondary teachers' knowledge and beliefs about subject matter and their impact on instruction. In J. Gess-Newsome and N.G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 51-94). Dordrecht: Kluwer Academic.

Gess-Newsome, J., Southerland, S. A., Johnson, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal, 40*, 731-767.

Gnadinger, C. (2008). Peer-mediated instruction: Assisted performance in the primary classroom. *Teachers and Teaching: Theory and Practice, 14,* 129-142.

Goldstein, H. (1995). *Multilevel statistical models.* London: Arnold.

*Goldston, M. J., Dantzler, J., Day, J., & Webb, B. (2013). A psychometric approach to the development of a 5E lesson plan scoring instrument for inquiry-based teaching. *Journal of Science Teacher Education, 24,* 527-551.

Goldston, M. J., Day, J. B., Sundberg, C., & Dantzler, J. (2010). Psychometric analysis of a 5E Learning Cycle Lesson Plan Assessment Instrument. *International Journal of Science and Mathematics Education, 8,* 633-648.

Goodrum, D., Hackling, M., & Rennie, L. (2001). The status and quality of teaching and learning of science in Australian schools. A research report for the Department of Education, Training and Youth Affairs. Canberra: Australia.

Goossens, E., Moens, L., Van Droogenbroeck, I., Verschueren, J., Smits, D., Van Opstal, M., Maes, F., D'Haenens, E., & Van Damme, J. (2009). *Observaties vierde leerjaar. Instrumentontwikkeling en basisrapportage (schooljaar 2006-2007).* [*Observations 4th grade. Instrument development and basic report (school year 2006-2007)]* Leuven, Belgium: Steunpunt 'Loopbanen doorheen Onderwijs naar Arbeidsmarkt', Cel 'Schoolloopbanen in het basisonderwijs' (SiBO).

Gott, R., & Duggan, S. (1996). Practical work: Its role in the understanding of evidence in science. *International Journal of Science Education, 18,* 791-806.

Goulart, M. I. M., & Roth, W.-M. (2010). Engaging young children in collective curriculum design. *Cultural Studies of Science Education, 5,* 533-562.

Gouwy, A., Plessers, N. & Willems, I. (2002). Leerkrachtstijl in de context van het kleuter en lager onderwijs: optimalisering van de Adult Style Observation Schedule. [Teacher style in the context of kindergarten and primary school: optimization of the Adult Style Observation Schedule], Unpublished Master's thesis, University of Leuven, Faculty of Psychology and Educational Sciences, Belgium.

Gustafson, B. J. & Rowell, P. M. (1995). Elementary preservice teachers: Constructing conceptions about learning science, teaching science and the nature of science. *International Journal of Science Education*, *17*, 589-605.

Hackling, M., Smith, P., & Murcia, K. (2011). Enhancing classroom discourse in primary science: The Puppets Project. *Teaching Science*, *57*, 18-25.

Hafen, C. A., Bridget, K. H., Hamre, J. P. A., Bell, C. A., Gitomer, D. H., Pianta, R. C. (2015). Teaching through interactions in secondary school classrooms: Revisiting the factor structure and practical application of the Classroom Assessment Scoring System – Secondary. *The Journal of Early Adolescence*, *35* (5-6), 651-680.

Hakkarainen, K. (2009). A knowledge-practice perspective on technology-mediated learning. *International Journal of Computer Supported Collaborative Learning, 4,* 213-231.

Haladyna, T., Olsen, R., and Shaughnesssy, J. (1982). Relations of student, teacher, and learning environment variables to attitudes to science. *Science Education*, *66*, 671-687.

Hamilton, J. (2003). Interaction, dialogue and a creative spirit of enquiry? In E. W. L. Norman, D. Spendlove, P. Grover, & A. Mitchell (Eds.), *DATA international research conference 2004* (pp. 35-44). Sheffield: Sheffield Hallam University.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76,* 949- 967.

Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early primary classrooms. In R.C. Pianta, M.J. Cox, & K. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 49-83). Baltimore, MD: Paul H. Brookes.

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). *Building a science of classrooms: Application of the CLASS framework in over 4,000 U.S. early childhood and primary classrooms.* New York: Foundation for Child Development. Retrieved from http://www.fed-us.org/resources/resources_show.htm?doc_id=507559.

Haney, J. J., Czerniak, C. M., & Lumpe, A. T. (1996). Teacher beliefs and intentions regarding the implementation of science education reform strands. *Journal of Research in Science Teaching, 33,* 971-993.

Haney, J. J., & Lumpe, A. T. (1995). A teacher professional development framework guided by reform policies, teachers' needs, and research. Journal of Science Teacher Education, 6, 187-196.

Haney, J. J., Lumpe, A. T., Czerniak, C. M., & Egan, V. (2002). From beliefs to actions: The beliefs and actions of teachers implementing change. *Journal of Science Teacher Education, 13,* 171-187.

Hansen, R. (1997). The value of a utilitarian curriculum: The case of technological education. *International Journal of Technology and Design Education, 7,* 111-119.

Hansen, P. J. K. (2010). An effective introduction to technology and design in Norwegian primary education. *Design and Technology Education: An International Journal, 15,* 58-67.

Harel, I., & Papert, S. (Eds.). (1991). *Constructionism* (pp. 141-150). Norwood, NJ: Ablex.

Harlen, W., & Holroyd, C. (1997). Primary teachers' understanding of concepts of science: Impact on confidence and teaching. *International Journal of Science Education, 19,* 93-105.

Harlen, W., & Léna, P. (2011). Introduction to the theme. In de Vries, M. J., van Keulen, H., Peters, S., & Walma van der Molen, J. (Eds.), *Professional development for primary teachers in science and technology. The Dutch VTB-Pro project in an international perspective* (pp. 1-305). Rotterdam: SensePublishers. Retrieved from http://link.springer.com.kuleuven.ezproxy.kuleuven.be/book/10.1007/978-94-6091-713-4/page/1

Hashweh, M. Z. (1987). Effects of subject matter knowledge in the teaching of biology and physics. *Teaching and Teacher Education, 3*, 109-120.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16,* 235-266.

Hodson, D. (2008). Towards research-based practice in the teaching laboratory. *Studies in Science Education, 41,* 167-177.

Hollon, R., Anderson, C., & Smith, E. (1980). *A system for observing and analyzing elementary school science teaching: A user's manual.* East Lansing: Michigan State University, Institute for Research on Teaching.

Holubova, R. (2008). Effective teaching methods project-based learning in physics. *US-China Educational Review, 5,* 27-36.

*Houston, L. S., Fraser, B. J., & Ledbetter, C. E. (2008). An evaluation of elementary school science kits in terms of classroom environment and student attitudes. *Journal of Elementary Science Education, 20,* 29-47.

Howes, C. (2000). Social-emotional classroom climate in child care, child-teacher relationships and children's second grade peer relations. *Social Development, 9*, 191-204.

Horizon Research (2000). *Inside the classroom observation and analytic protocol.* Chapel Hill, NC: Horizon Research.

Horizon Research, Inc. (2003). 2005-06 Data Collection Manual. Retrieved from http://www.horizon-research.com/LSC/ manual/

*Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston, K., ...

Hugener, L., Pauli, C., Reusser, K., Lipowsky, E., Rakoczy, K., & Klieme, E. (2009). Teaching patterns and learning quality in Swiss and German mathematics lessons. *Learning and Instruction, 19,* 66-78.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes, and efficacy. *Education Policy Analysis Archives, 13,* 1-26.

Jacobowitz, R. (1997). 30 tips for effective teaching. *Science Scope, 21,* 22-5.

James, M. C., & Scharmann, L. C. (2007). Using analogies to improve the teaching performance of preservice teachers. *Journal of Research in Science Teaching, 44,* 565-585.

Jarvis, T., & Pell, A. (2004). Primary teachers' changing attitudes and cognition during a two-year science inservice program and their effect on pupils. *International Journal of Science Education, 26,* 1787-1811.

Jarvis, T., & Pell, A. (2005). Factors influencing primary school children's attitudes toward science before, during, and after a visit to the UK National Space Centre. *Journal of Research in Science Teaching, 42,* 53-83.

Johnston, J., & Ahtee, M. (2006). Comparing primary student teachers' attitudes, subject knowledge and pedagogical content knowledge needs in a physics activity. *Teaching and Teacher Education, 22,* 503-512.

*Johnson, B., & McClure, R. (2004). Validity and reliability of a shortened, revised version of the Constructivist Learning Environment Survey (CLES). *Learning Environments Research, 7,* 65-80.

Jones, M. G., & Carter, G. (2007). Science teacher attitudes and beliefs. In Abell, S. K., Lederman, N. G. (Eds.), *Handbook of research on science education (pp. 1067-1104).* Mahwah, NJ: Lawrence Erlbaum Associates.

Jones, A., Buntting, C., & de Vries, M. J. (2013). The developing field of technology education: A review to look forward. *International Journal of Technology and Design Education, 23,* 191-212.

Jones, B. F., Rasmussen, C. M., & Moffitt, M. C. (1997). *Real-life problem solving: A collaborative approach to interdisciplinary learning.* Washington, DC: American Psychological Association.

Jorde, D., & Dillon, J. (2013). Science education research and practice in Europe: retrospective and prospective. The Netherlands: Rotterdam, Sense Publishers.

Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction, 23,* 33-42.

Kafai, Y., & Resnick, M. (Eds.) (1996). *Constructionism in practice: Designing,* *thinking and learning in a digital world.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kaldi, S., Filippatou, D., & Govaris, C. (2011). Project-based learning in primary schools: Effects on pupils' learning and attitudes. Education 3-13: *International Journal of Primary, Elementary and Early Years Education, 39*, 35-47.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teachers: Combining high-quality observations with students surveys and achievement gains.* Policy and practice brief prepared for the Bill and Melinda Gates Foundation.

Kang, N. H. (2007). Elementary teachers' epistemological and ontological understanding of teaching for conceptual learning. *Journal of Research in Science Teaching, 44*, 1292-1317.

Katz, L., Sadler, K., & Craig, D. V. (2005). Science professors serve as mentors for early childhood preservice teachers in the design and implementation of standards-based science units. *Journal of Elementary Science Education, 17*, 43-55.

Katz, D., & Stotland, E. (1959). A preliminary statement to a theory of attitude structure and change. In Ostrom, T.M.: The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology, 5*, 12-30.

Kaya, K. (2014). Dynamic variables of science classroom discourse in relation to teachers' instructional beliefs. *Australian Journal of Teacher Education, 39*, 57-74.

Keys, C. W., & Bryan, L. A. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching, 38,* 631-645.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41,* 75-86.

Klop, T., & Severiens, S. (2007). An exploration of attitudes towards modern biotechnology: A study among Dutch secondary school students. *International Journal of Science Education, 29,* 663-679.

Koballa, T. R., Jr., & Crawley, F. E. (1985). The influence of attitude on science teaching and learning. *School Science and Mathematics, 85,* 222-232.

Koch, J., & Appleton, K. (2007). The effect of a mentoring model for elementary science professional development. *Journal of Science Teacher Education, 18,* 209-231.

Kolodner, J. L. (2001). A note from the editor. *Journal of the Learning Sciences, 10,* 1-4.

Kolodner, J. L. (2006a). Case-based reasoning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 225-242). Cambridge: Cambridge University Press.

Kolodner, J. L. (2006b). A note from the editor. *Journal of the Learning Sciences, 15,* 1-2.

Kotelawala, U. (2012). Lesson study in a methods course: Connecting teacher education to the field. *The Teacher Educator, 47,* 67-89.

Koutsides, G. (2001). Using cooperative learning in design and technology. *The Journal of Design and Technology Education, 6,* 55-59.

Kozma, R., Belzer, S., & Jaffe, J. M. (1993). *BioMap: An interactive hypermedia environment to promote conceptual understanding in biology.* Paper presented at American educational Research Association, Atlanta, GA.

Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences, 7,* 313-350.

Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The Elementary School Journal, 94,* 483-497.

Kyle, W. C. (1994). School reform and the reform of teacher education: Can we orchestrate harmony? *Journal of Research in Science Teaching, 31,* 785-786.

Laevers, F. (1998). Understanding the world of objects and of people: Intuition as the core element of deep level learning. *International Journal of Educational Research, 29,* 69-86.

Laevers, F. (1999). The project Experiential Education. Well-being and involvement make the difference. *Early Education, 27.*

Laevers, F. (2000). Forward to basics! Deep-level-learning and the experiential approach. *Early Years, 20,* 20-29.

Laevers, F. (2005). *Deep-level-learning and the experiential approach in early childhood and primary education.* Leuven: Katholieke Universiteit Leuven, Research Centre for Early Childhood and Primary Education.

Laevers, F. (2006). Onderwijs op een nieuwe leest. Een reconstructie van de actuele paradigmashift [A new approach for education. A reconstruction of the actual paradigmshift]. In: F. Laevers, and E. Bertrands (Eds.,), Draagkracht geven. Uitkomst voor de toekomst. (pp. 5-24). Leuven: CEGO.

Laevers, F. (2011). Werkzame bestanddelen van een krachtige leeromgeving voor techniek [Active constituents of a powerful learning environment for technology]. In F. Laevers, & L. Heylen (Eds.), *Passie voor wetenschap en techniek. Onderzoekend en ontwerpend leren in de basisschool* [Passion for science and technology. Inquiry and design learning in primary school] (pp. 44-64). Averbode: CEGO.

Laevers, F., Berghmans, I., Declercq, B., & Buyse, E. (2015). Intensive Seminar. A process-oriented approach to quality in early childhood education. Assessment tools for research [Hand-outs].

Laevers, F., Declercq, B., & Jackaman, S. (2011). Observing involvement in children from 6 to 12 years. Centre for Experiential Education and Kent County Council [DVD Training Pack]. Leuven: CEGO Publishers.

Laevers, F., & Heylen, L. (2003). *Involvement of children and teacher style. Insights from an international study on experiential education.* Studia Paedagogica 35. Leuven: Leuven University Press.

Laevers, F., & Heylen, L. (2013). *Een procesgerichte aanpak voor 6- tot 12-jarigen in het basisonderwijs* [A process-oriented approach for 6- to 12-year olds in primary education]. Leuven: CEGO.

Laevers, F., Heylen, L., & Daniëls, D. (2004). *Ervaringsgericht werken met 6- tot 12 jarigen in het Basisonderwijs* [Working experientially with 6- tot 12- year olds in Primary school]. Leuven, Belgium: CEGO Publishers.

Laevers, F., & Laurijssen, J. (2001). Well-being, involvement and satisfaction of pupils in nursery and primary school: A guide to systematic observation and survey [Internal report]. *Leuven: KU Leuven, Center for Nursery and Primary Education.*

*Lakshmanan, A., Heath, B. P., Perlmutter, A., & Elder, M. (2011). The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction. *Journal of Research in Science Teaching, 48,* 534-551.

Lawrenz, F., Huffman, D., & Appeldorn, K. (2002). *CETP core evaluation: Classroom observation handbook.* Minneapolis-St. Paul, MN: University of Minnesota.

Lawson, A. E. (1995). *Science teaching and the development of thinking.* Belmont, Ca: Wadsworth.

Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student engagement research, policy, and practice. *Review of Educational Research, 83,* 432-479.

Lee, O., Deaktor, R. A., Hart, J. E., Cuevas, P., & Enders, C. (2005). An instructional intervention's impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching, 42,* 857-887.

*Lee, O., Hart, J. E., Cuevas, P., & Enders, C. (2004). Professional development in inquiry-based science for elementary teachers of diverse student groups. *Journal of Research in Science Teaching, 41,* 1021-1043.

*Lee, O., Lewis, S., Adamson, K., Maerten-Rivera, J., & Secada, G. W. (2007). Urban elementary school teachers' knowledge and practices in teaching science to English language learners. *Science Teacher Education, 92,* 733-758.

*Lee, O., Luykx, A., Buxton, C., & Shaver, A. (2007). The challenge of altering elementary school teachers' beliefs and practices regarding linguistic and cultural diversity in science instruction. *Journal of Research in Science Teaching, 44,* 1269-1291.

Lento, C. (in preparation). *Leerkrachtstijl in de context van het lager onderwijs: Optimalisering van de Adult Style Observation Schedule* [Teacher style in the context of primary education: Optimization of the Adult Style Observation Schedule]. Unpublished Master's thesis, University of Leuven, Faculty of Psychology and Educational Sciences, Belgium.

*Levy, J. A., Pasquale, M. M., & Marco, L. (2008). Models of providing science instruction in the elementary grades: A research agenda to inform decision makers. *Science Educator, 17,* 1-18.

Liang, L., & Gabel, D. (2005). Effectiveness of a constructivist approach to science instruction for prospective elementary teachers. *International Journal of Science Education, 27,* 1143-1162.

Liang, L. L., & Richardson, G. M. (2009). Enhancing prospective teachers' science teaching efficacy beliefs through scaffolded, student-directed inquiry. *Journal of Elementary Science Education, 21,* 51-66.

Liljeström, A., Enkenberg, J., & Pöllänen, S. (2013). Making learning whole: An instructional approach for mediating the practices of authentic science inquiries. *Cultural Studies of Science Education, 8,* 51-86.

Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading and Writing Quarterly, 19,* 119-137.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction, 19,* 527-537.

Liu, X. (2010). *Using and developing measurement instruments in science education. A*

Rasch modeling approach. Science and Engineering Education Sources. Concordia University, Charlotte, North Carolina: Information Age Publishing.

Liu, X. (2012). Developing measurement instruments for science education. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (Chapter 43, 651-665). Dordrecht: Springer. Retrieved from http://link.springer.com/chapter/10.1007%2F978-1-4020-9041-7_43

Liu, M., & Hsiao, Y. (2002). Middle school students as multimedia designers: A project-based learning approach. *Journal of Interactive Learning Research, 13,* 311-337.

Loughran, J., Mulhall, P., & Berry, A. (2008). Exploring pedagogical content knowledge in science teacher education. *International Journal of Science Education, 30,* 1301-1320.

Luckner, A. E., & Pianta, R. C. (2011). Teacher-student interactions in fifth grade classrooms: Relations with children's peer behavior. *Journal of Applied Developmental Psychology, 32,* 257-266.

Luera, G. R., & Otto, C. A. (2005). Development and evaluation of an inquiry-based elementary science teacher education program reflecting current reform movements. *Journal of Science Teacher Education, 16,* 241-258.

*Luykx, A., & Lee, O. (2007). Measuring instructional congruence in elementary science classrooms: Pedagogical and methodological components of a theoretical framework. *Journal of Research in Science Teaching, 44,* 424-447.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome, & N.G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95-132). Dordrecht, The Netherlands: Kluwer.

Mansour, N. (2013). Modelling the sociocultural contexts of science education: The teachers' perspective. *Research in Science Education, 43,* 347-369.

Mant, J., Wilson, H., & Coates, D. (2007). The effect of increasing conceptual challenge in primary science lessons on pupils' achievement and engagement. *International Journal of Science Education, 29,* 1707-1719.

Marks, H. (2000). Student engagement in instructional activity: Patterns in the primary, middle, and high school years. *American Educational Research Journal, 37,* 153-184.

*Marshall, C. J., Horton, R., Igo, L. B., & Switzer, M. D. (2007). K-12 science and mathematics teachers' beliefs about and use of inquiry in the classroom. *International Journal of Science and Mathematics Education, 7,* 575-596.

Martin-Dunlop, C. & Fraser, B. J. (2008). Learning environment and attitudes associated with an innovative science course designed for prospective elementary teachers. *International Journal of Science and Math Education, 6,* 163-190.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Blunk, M., Crawford, B., Kelly, B., & Meyer, K. M. (1994). Enacting project-based science: Experiences of four middle grade teachers. *The Elementary School Journal, 94,* 517-538.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., & Soloway, E. (1997). Enacting project-based science. *The Elementary School Journal, 97,* 341-358.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., & Soloway, E. (1998). New technologies for teacher professional development. *Teaching and Teacher Education, 14,* 33-52.

McDevitt, T. M., Heikkinen, H. W., Alcorn, J. K., Ambrosio, A. L., & Gardner, A. L. (1993). Evaluation of the preparation of teachers in science and mathematics: Assessment of preservice teachers' attitudes and beliefs. *Science Education, 77,* 593-610.

McDuffie, T. E. Jr. (2001). Scientists – geeks and nerds? *Science and Children, 38,* 16-19.

Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. The *British Journal of Educational Psychology, 80,* 1-14.

Mercer, N., Dawes, L., & Staarman, J. K. (2009). Dialogic teaching in the primary science classroom. *Language and Education, 23,* 353-369.

Meyer, X., & Crawford, B. A. (2011). Teaching science as a cultural way of knowing: Merging authentic inquiry, nature of science, and multicultural strategies. *Cultural Studies of Science Education, 6,* 525-547.

Meyer, D., & Turner, J. (2002). Discovering emotion in classroom motivational research. *Educational Psychologist, 37,* 107-114.

Myers, R. E. & Fouts, J. T. (1992). A cluster analysis of high school science learning environments and attitude toward science. *Journal of Research in Science Teaching, 29*, 929-937.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Mitra, S. (2014). The future of schooling: Children and learning at the edge of chaos. *Prospects, 44*, 547-558.

Mitra, S., & Dangwal, R. (2010). Limits to self-organising systems of learning – the Kalikuppam experiment. *British Journal of Educational Technology, 41*, 672-688.

Moos, R. H. (1979). *Evaluating educational environments.* San Francisco, CA: Jossey-Bass.

Morgan, & Slough, (2013). Classroom management considerations: Implementing STEM project-based learning. In: Capraro, R.M., Capraro, M.M., & James, R.M. (2013) (Eds.), *STEM project-based learning: An integrated Science, Technology, Engineering, and Mathematics (STEM) approach* (Chapter 11, 99-107). Sense Publishers: Rotterdam.

Mueller, M. P. (2011). Ecojustice in science education: Leaving the classroom. *Cultural Studies of Science Education, 6*, 351-360.

Mulder, M. (2001). Competentieontwikkeling in organisaties. *Perspectieven en praktijk.* [Competence development in organisations. Perspectives and practices]. 's-Gravenhage: Elsevier Bedrijfs Informatie.

Munck, M. (2007). Science pedagogy, teacher attitudes, and student success. *Journal of Elementary Science Education, 19,* 13-24.

Murphy, C., & Beggs, J. (2003). Children's perceptions of school science. *School Science Review, 84,* 109-16.

Murphy, C., & Smith, G. (2012). The impact of a curriculum course on pre-service primary teachers' science content knowledge and attitudes towards teaching science. *Irish Educational Studies, 31,* 77-95.

National Institute of Child Health and Human Development, Early Child Care Research Network (2002). The relation of first grade classroom environment to structural classroom features, teacher, and student behaviors. *The Primary School Journal, 102,* 367-387.

National Institute of Child Health and Human Development, Early Child Care Research Network (2003). Social functioning in first grade: Prediction from home, child care and concurrent school experience. *Child Development, 74,* 1639-1662.

National Institute of Child Health and Human Development (NICHD), Early Child Care Research Network (2006). *The relations of classroom contexts in the early elementary years to children's classroom and social behavior.* In A. Huston & M.N. Ripke (Eds.), Developmental contexts in middle childhood (pp. 217-236). New York, NY: Cambridge University Press.

National Research Council (1996). *National science education standards.* Washington, DC: National Academy Press.

*Neale, D. C., Smith, D., & Johnson, V. G. (1990). Implementing conceptual change teaching in primary science. *The Elementary School Journal, 91,* 109-131.

Newmann, F.M., & Associates (1996). *Authentic achievement: Restructuring schools for intellectual quality.* San Francisco: Jossey-Bass.

*Newman, D., Finney, P. B., Bell, S., Turner, H., Jaciw, A. P., Zacamy, J. L., & Gould, L. F. (2012). *Evaluation of the effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI).* Final report. NCEE 2012-4008.

Newmann, F.M., Secada, W.G., & Wehlage, G.G. (1995). *A guide to authentic instruction and assessment: Vision, standards and scoring.* Madison, WI: University of Wisconsin-Madison, Center on Organization and Restructuring of Schools, Wisconsin Center for Education Research.

Newmann, F., Wehlage, G., & Lamborn, S. (1992). The significance and sources of student engagement. In F. Newman (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11-39). New York: NY: Teachers' College Press.

Niemiec, C.P., & Ryan, R.M. (2009). Autonomy, competence and relatedness in the classroom. Applying self-determination theory to education practice. *Theory and Research in Education, 7,* 133-144.

Nilsson, P., & Loughran, J. J. (2012). Understanding and assessing primary science student teachers pedagogical content knowledge. *Journal of Science Teacher Education, 23,* 699-721.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26,* 327-257.

Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes, 35,* 135-198.

Oberon (2009). Een selectie uit de tussenrapportage effectstudie VTB-Pro. Schooljaar 2008-2009. [A selection of the interim report of the outcome study VTB-Pro. School year 2008-2009.] Utrecht: Platform Bèta Techniek.

Oberon (2011). Effectstudie VTB-Pro. Eindrapport. [Effect study VTB-Pro. Final Report.] Utrecht: Platform Bèta Techniek.

Organisation for Economic Co-operation and Development (2007). *PISA 2006: Science competencies for tomorrow's world.* Paris: OECD.

Osborne, J.F., & Collins, S. (2000). Pupils' and parents' views of the school science curriculum. *School Science Review*, 82, 23(9).

Osborne, J., & Dillon, J. (2008). *Science education in Europe: Critical reflections (a report to the Nuffield Foundation).* London: the Nuffield Foundation. Retrieved from http://www.pollen-europa.net/pollendev/ImagesEditor/Nuffieldreport.pdf.

Osborne, J., Simon , S., & Collins, S. (2003). Attitudes towards science: a review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049-1079.

Osborn, M., McNess, E., Broadfoot, P., with Pollard, A., & Triggs, P. (2000). *What teachers do: Changing policy and practice in primary education.* London and New York: Continuum.

Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: a study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching, 50,* 315-347.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 307-332.

Palinscar, A. S., Anderson, C. A., & David, Y. M. (1993). Pursuing scientific literacy in the middle grades through collaborative problem solving. *Primary School Journal, 93,* 643-658.

Palmer, D. H. (2001). Factors contributing to attitude exchange amongst preservice elementary teachers. *Science Education, 86,* 122-138.

Palmer, D. H. (2004). Situational interest and the attitude towards science of primary teacher education students. *International Journal of Science Education*, 26, 895-908.

Papert, S. (1980). *Mindstorms, children, computers and powerful ideas.* New York: Basic Books Inc.

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38,* 261-284.

Parker, J. & Spink, E. (1997). Becoming science teachers: An evaluation of the initial stages of primary teacher training. *Assessment & Evaluation in Higher Education, 22,*17-31.

Patrick, B., Hisley, J., & Kempler, T. (2000). What's everybody so excited about? The effects of teacher enthusiasm on student instrinsic motivation and vitality. *The Journal of Experimental Education, 68,* 217-236.

Paulson, S.E., Marchant, G.J., & Rothlisberg, B.A. (1998). Early adolescents' perceptions of patterns of parenting, teaching, and school atmosphere: Implications for achievement. *The Journal of Early Adolescence, 18,* 5-26.

Pehmer , A.-K., Gröschner, A., & Seidel, T. (2015). How teacher professional development regarding classroom dialogue affects students' higher-order learning. *Teaching and Teacher Education, 47,* 108-119.

Pell, T., & Jarvis, T. (2001). Developing attitude to science scales for use with children of ages from five to eleven years. *International Journal of Science Education, 23,* 847-862.

Petticrew, M., & Roberts, H. (2006). *Systemic reviews in the social sciences: A practical guide.* Oxford: Blackwell.

Pettigrew, J., Miller-Day, M., Shin, Y., Hecht, M.L., Krieger, J.L., & Graham, J.W. (2013). Describing teacher-student interactions: a qualitative assessment of teacher implementation of the 7th grade keepin' it REAL substance use interverntion. *American Journal of Community Psychology, 51,* 43-56.

Piaget, J., & Inhelder, B. (1969). *The Psychology of the Child.* New York: Basic Books.

Pianta, R. C., Hamre, K. B., & Mintz, S. (2012). *Classroom Assessment Scoring System, Upper Primary Manual. Charlottesville: Teachstone.*

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS).* Baltimore, MD: Paul H. Brookes.

Pianta, R.C., Nimetz, S.L., & Bennett, E. (1997). Mother-child relationships, teacher-child relationships, and school outcomes in preschool and kindergarten. *Early Childhood Research Quarterly, 12,* 263-280.

Piburn, M., & Sawada, D. (2000). *Reformed teaching observation protocol (RTOP)* *reference manual.* Technical Report (ACEPT- TR-IN00-32). Arlington, VA: National Science Foundation.

Piburn, M., Sawada, D., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). *Reformed Teaching Observation Protocol (RTOP) Reference Manual* (ACEPT Technical Report No. IN00-3). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Pirozzo, R. (1987). A self-directed independent approach to learning science. *Gifted Child Today, 10,* 22-24.

Polman, J. L., & Pea, R. D. (2001). Transformative communication as a cultural tool for guiding inquiry science. *Science Education, 85,* 223-238.

Plonczak, I. (2008). Science for all: Empowering elementary school teachers. *Education, Citizenship and Social Justice, 3,* 167-181.

*Pop, M. M., Dixon, P., & Grove, C. M. (2010). Research experiences for teachers (RET): Motivation, expectations, and changes to teaching practices due to professional program involvement. *Journal of Science Teacher Education, 21,* 127-147.

Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards S&T at K-12 levels: a systematic review of 12 years of educational research. *Studies in Science Education, 50,* 85-129.

Pressley, M., & El-Dinary, P.B. (1993). Introduction. *Elementary School Journal, 94,* 105-08.

Pucel, D. (1992). *Technology education: In changing role within general education.* Paper presented at the American Vocational Association Convention, St. Louis, MO. Retrieved from http://files.eric.ed.gov/fulltext/ED353400.pdf

Rahm, J., Miller, H., Hartley, L., & Moore, J. (2003). The value of an emergent notion of authenticity: Examples from two student/teacher-scientist partnership programs. *Journal of Research in Science Teaching, 40,* 737-756.

Ramey-Gassert, L., Schroyer, M. G., & Staver, J. R. (1996). A qualitative study of factors influencing science teaching self-efficacy of primary level teachers. *Science Education, 80,* 283-315.

Rashbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2005). *MLwiNVersion 2.02.* Bristol: Centre for Multilevel Modelling, University of Bristol.

Raver, C.C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75,* 346-353.

Reinsvold, L. A., & Cochran, K.F. (2012). Power dynamics and questioning in primary science classrooms. *Journal of Science Teacher Education, 23,* 745-768.

*Rennie, L. J., Goodrum, D., & Hackling, M. (2001). Science teaching and learning in Australian schools: Results of a national study. *Research in Science Education, 31,* 455-498.

Rimm-Kaufman, S., La Paro, K., Downer, J., & Pianta, R. (2005). The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *The Primary School Journal, 105,* 377-394.

Rivet, A. E., & Krajcik, J. S. (2008). Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. *Journal of Research in Science Teaching, 45,* 79-100.

Roehrig, G. H., & Luft, J. A. (2004). Constraints experience by beginning secondary science teachers in implementing scientific inquiry lessons. *International Journal of Science Education, 26,* 3-24.

Rogers, C. (1979). Leren in vrijheid [Learning in freedom]. Haarlem, the Netherlands: De Toorts.

Rogers, M. A. P., Cross, D. I., Gresalfi, M. S., Trauth-Nare, A. E., & Buck, G. A. (2011). First year implementation of a project-based learning approach: The need for addressing teachers' orientations in the era of reform. *International Journal of Science and Mathematics Education, 9,* 893-917.

Rohaan, E. J., Taconis, R., & Jochems, W. M. G. (2010). Reviewing the relations between teachers' knowledge and pupils' attitudes in the field of primary technology education. *International Journal of Technology and Design Education, 20,* 15-26.

Rohaan, E. J., Taconis, R., & Jochems, M. G. W. (2012). Analysing teacher knowledge for technology education in primary schools. *International Journal of Design and Technology Education, 22,* 271-280.

Roorda, D. L., & Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research, 81*, 493-521.

Rosebery, A. S., Warren, B., & Conant, F. R. (1992). Appropriating scientific discourse: Findings from language minority classrooms. *Journal of the Learning Sciences, 2,* 61-94.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitude. In M. Rosenberg, C. Hovland, W.McGuire, R. Abelson, & J. Brehm (Eds.), *Attitude organization and change.* New Haven: CT: Yale University Press.

Ross, J. A. (1998). The antecedents and consequences of teacher efficacy. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 7, pp. 49-74). Greenwich, CN: JAL.

Rossman, A. D. (1993). Managing hands-on inquiry. *Science and Children, 31*, 35-37.

Roth, W. -M. (1998a). *Designing communities.* Boston: Kluwer.

Roth, W.-M. (1998b). Teacher questioning in an open-inquiry learning environment: Interactions of context, content, and student responses. *Journal of Research in Science Teaching, 33*, 709-736.

Roth, W.-M. (2001). Learning science through technological design. *Journal of Research in Science Teaching, 38,* 768-790.

*Roth, W.-M., & Bowen, G. M. (1995). Knowing and interacting: A study of culture, practices, and resources in a Grade 8 open-inquiry science classroom guided by a cognitive apprenticeship metaphor. *Cognition and Instruction, 13,* 73-128.

Ryan, R.M., & Connell, J.P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57,* 749-761.

Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55,* 68-78.

Ryan, R.M., & Powelson, C.L. (1991). Autonomy and relatedness as fundamental to motivation and education. *Journal of Experimental Education, 60,* 49-66.

Sahin, C., Bullock, K., & Stables, A. (2002). Teachers' beliefs and practices in relation to their beliefs about questioning at key stage 2. *Educational Studies, 28,* 371-384.

Savers, J. R., & Duffy, T. M. (1995). Problem-based learning: An instructional model and its constructive framework. *Educational Technology, 35,* 31-38.

Sawada, D., Piburn, M., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). Reformed Teaching Observation Protocol (RTOP) (ACEPT Technical Report No. IN00-1). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Sawyer, R. K. (2004). Creative teaching: Collaborative discussion as disciplined improvisation, *Educational Researcher, 33,* 12-20.

Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. *Cognition and Instruction, 9*, 177-199.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1*, 61-80.

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36,* 111-139.

Schweinhart, L. J., & Weikart, D. P. (1997). The high/scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly, 12,* 117-143.

Scott, P. H., Mortimer, E. F., & Aguiar, O. G. (2006). The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education, 90*, 605-631.

Shepardson, D. P., & Pizzini, E. L. (1992). Gender bias in female elementary teachers' perceptions of the scientific ability of students. *Science Education, 76,* 147-153.

Shrigley, R. L. (1983). Persuade, mandate, and reward: A paradigm for changing the science attitudes and behaviors of teachers. *School Science and Mathematics, 83,* 204-215.

Shulman, (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15,* 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57,* 1-23.

Shymansky, A. J., Wang, T.-Z., Annetta, A. L., Yore, D. L., & Everett, A. S. (2010). How much professional development is needed to effect positive gains in K-6 student achievement on high stakes science tests. *International Journal of Science and Mathematics Education, 10,* 1-19.

Sierens, E., Soenens, B., Vansteenkiste, M., Goossens, L., & Dochy, F. (2006). De authoritatieve leerkrachtstijl: Een model voor de studie van leerkrachtstijlen [The authoritative teacher style. A model for the study of teacher styles.]. *Pedagogische Studiën, 83,* 419-431.

Sierens, E., Vansteenkiste, M., Goossens, L., Soenens, B., & Dochy, F. (2009). The synergistic relationships of perceived autonomy support and structure in the prediction of self-regulated learning. *British Journal of Educational Psychology, 79,* 57-68.

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42,* 70-83.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis.* New York: Oxford University Press.

Skager, R. (1984). *Organizing schools to encourage self-direction in learners.* Oxford: Pergamon.

Skamp, K. & Mueller, A. (2001). Student teachers' conceptions about effective primary science teaching: A longitudinal study. *International Journal of Science Education, 23,* 331-351.

Skinner, E.A., & Belmont, M.J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85,* 571-581.

Soenens, B., Sierens, E., Vansteenkiste, M., Dochy, F., & Goossens, L. (2012). Psychologically controlling teaching: Examining outcomes, antecedents, and mediators. *Journal of Educational Psychology, 104,* 108-120.

Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching, 51,* 870-901.

Smart, J. B., & Marshall, J. C. (2013). Interactions between classroom discourse, teacher questioning, and student cognitive engagement in middle school science. *Journal of Science Teacher Education, 24,* 249-267.

Smeets, A. (2014). Onderzoek naar de parallelliteit van twee gestandaardiseerde situaties voor de meting van leerkrachtstijl in de context van wetenschap en techniek activiteiten [Investigation of the analogy of two standardized situations for the measurement of teacher style in the context of S&T activities]. Unpublished Master's thesis, University of Leuven, Faculty of Psychology and Educational Sciences, Belgium.

Smith, G. (2014). An innovative model of professional development to enhance the teaching and learning of primary science in Irish schools. *Professional Development in Education, 40,* 467-487.

Smith, G. (2015). The impact of a professional development programme on primary teachers' classroom practice and pupils' attitudes to science. *Research in Science Education, 45,* 215-239.

Smith, E. L., Blakeslee, T. D., & Anderson, C. W. (1993). Teaching strategies associated with conceptual change learning in science. *Journal of Research in Science Teaching, 30,* 111-126.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling.* London: Sage.

Soares, D.A., & Vannest, K.J. (2013). STEM project-based learning and teaching for exceptional learners. In: Capraro, R.M., Capraro, M.M., & James, R.M. (2013) (Eds.), *STEM project-based learning: An integrated Science, Technology, Engineering, and Mathematics (STEM) approach* (Chapter 10, 85-98). Sense Publishers: Rotterdam. Retrieved from http://link.springer.com.kuleuven.ezproxy.kuleuven.be/book/10.1007/978-94-6209-143-6/page/1

Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychologist, 39*, 97-110.

Stevens, C., & Wenner, G. (1996). Elementary preservice teachers' knowledge and beliefs regarding science and mathematics. *School Science and Mathematics, 96*, p. 2 (8).

Stuart, C., & Thurlow, D. (2000). Making it their own: Preservice teachers' experiences, beliefs, and classroom practices. *Journal of Teacher Education, 51*, 113-121.

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching, 49*, 515-537.

Sweertvaegher, S. (2008). De relatie tussen de psychologische basisbehoeften en leerkrachtstijl. Speelt autonome motivatie een mediërende rol? [The connection between the psychological basic needs and teacher style. Does autonomous motivation play a role?]. Unpublished Master's thesis, University of Leuven, Faculty of Psychology and Educational Sciences, Belgium.

Tabak, I., & Baumgartner, E. (2004). The teacher as partner: Exploring participant structures, symmetry, and identity work in scaffolding. *Cognition and Instruction, 22*, 393-429.

Talton, E. L., & Simpson, R. D. (1987). Relationships of attitude toward learning environment with attitude toward and achievement in science among tenth grade biology students. *Journal of Research in Science Teaching, 24*, 507-525.

Tamir, P., Stavy, R., & Ratner, N. (1998). Teaching science by inquiry: Assessment and learning. *Journal of Biological Education, 33,* 27-32.

*Tao, Y., Oliver, M., & Venville, G. (2013). A comparison of approaches to the teaching and learning of science in Chinese and Australian elementary classrooms: Cultural and socioeconomic complexities. *Journal of Research in Science Teaching, 50,* 33-61.

Taylor, B.M., Pearson, P.D., Peterson, D.S., & Rodriguez, M.C. (2005). The CIERA School Change Framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly, 40,* 40-69.

Taylor, P., Dawson, V., & Fraser, B. (1995). *A constructivist perspective on monitoring classroom learning environments under transformation.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

*Taylor, P. C., & Fraser, B. J. (1991). *CLES: An instrument for assessing constructivist learning environments.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Lake Geneva, WI.

Taylor, P., Fraser, B., & Fisher, D. (1993). *Monitoring the development of constructivist learning environments.* Paper presented at the annual convention of the National Science Teachers Association, Kansas City, MO.

Taylor, P., Fraser, B., & Fisher, D. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research, 27,* 293-302.

Thomas, J. W. (2000). *A review of research on project-based learning.* Report prepared for the Autodesk Foundation. Retrieved from http://www.bie.org/files/researchreviewPBL.pdf

Thomas, J., Pedersen, J., & Finson, K. (2001). Validating the draw-a-science teacher-test checklist: Exploring mental models and teacher beliefs. *Journal of Science Teacher Education, 12,* 295-310.

Thompson, C. L., & Shrigley, R. L. (1986). What research says: Revising the science attitude scale. *School Science and Mathematics, 86,* 331-343.

Trumbull, D., Scarano, G., & Bonney, R. (2006). Relations among two teachers' practices and beliefs, conceptualizations of the nature of science, and their implementation of student independent inquiry projects. *International Journal of Science Education, 28,* 1717-1750.

Tucker, C.M., Zayco, R.A., Herman, K.C., Reinke, W.M., Trujillo, M., Carraway, K., Wallack, C., & Ivery, P.D. (2002). Teacher and child variables as predictors of academic engagement among low-income African American children. *Psychology in the Schools, 39,* 477-488.

Trumper, R. (1998). The need for change in elementary-school teacher training: The force concept as an example. *Asia-Pacific Journal of Teacher Education, 26,* 7-25.

*Türkmen, H. (2009). An effect of technology based inquiry approach on the learning of 'Earth, Sun, & Moon' subject. *Asia-Pacific Forum on Science Learning and Teaching, 10.*

.

Van Aalderen-Smeets, S., & Walma van der Molen, J.H.W. (2013). Measuring primary teachers' attitudes toward teaching science: Development of the Dimensions of Attitude Toward Science (DAS) Instrument. *International Journal of Science Education, 35,* 577-600.

Van Aalderen-Smeets, S. I. & Walma Van der Molen , J. H. (2015). Improving primary teachers' attitudes toward science by attitude-focused professional development. *Journal of Research in Science Teaching, 52,* 710-734.

Van Aalderen-Smeets, S. I., Walma van der Molen, J. H., & Asma, J. F. L. (2011). Primary teachers' attitudes towards science: A new theoretical framework. *Science Education, 96,* 158-182.

Van Cleynenbreugel, C., De Winter, V., Buyse, E., & Laevers, F. (2011). Understanding the physical world: Teacher and pupil attitudes towards science and technology.In: de Vries, M.J., van Keulen, H., Peters, S., & van der Molen, J.W. (Eds.), *Professional development of primary teachers in science and technology. The Dutch VTB-Pro project in an international perspective* (pp.121-143). Rotterdam: SensePublishers.

Van Driel, J. (2010). *Model-based development of science teachers' pedagogical content knowledge.* Paper presented at the International Seminar on Professional Reflections, National Science Learning Centre, UK.

Van Driel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. *Journal of Research in Science Teaching, 38,* 137-158.

Van Graft, M., & Kemmers, P. (2007). Onderzoekend en ontwerpend leren bij natuur en techniek: Basisdocument over de didactiek voor onderzoekend en ontwerpend leren in het primair onderwijs [Inquiry and design learning in nature and technology: Basic document about the didactics for inquiry and design learning in primary school]. Den Haag: Stichting Platform Bèta Techniek.

Van Heddegem, I., Gadeyne, E., Vandenberghe, N., Laevers, F., & Van Damme, J. (2004). Longitudinaal onderzoek in het basisonderwijs. Observatie-instrument schooljaar 2002-2003 (LOA-rapport nr. 20) [Longitudinal research in primary school: Observational measure school year 2002-2003 (LOA report no. 20)]. Leuven: Steunpunt 'Loopbanen doorheen Onderwijs naar Arbeidsmarkt', Cel 'Schoolloopbanen in het basisonderwijs' (SiBO).

Vallerand, R.J., Fortier, M.S., & Guay, F. (1997). Self-determination and persistence in real-life setting: Toward a motivational model of high school dropout. *Journal of Personality and Social Psychology, 72,* 1161-1176.

Van Droogenbroeck, I., Joosten, R., Imberechts, H., & Van Damme, J. (2010). *Longitudinaal onderzoek in het basisonderwijs. Observaties vijfde leerjaar: instrumentontwikkeling en basisrapportage (schooljaar 2007-2008).* [Longitudinal research in primary school. Observations 5th grade: instrument development and basic report (school year 2007-2008)] Leuven, België: Steunpunt 'Loopbanen doorheen Onderwijs naar Arbeidsmarkt', Cel 'Schoolloopbanen in het basisonderwijs' (SiBO).

Vansteenkiste, M., Lens, W., & Deci, E.L. (2006). Intrinsic versus extrinsic goal-contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist, 41*, 19-31.

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology, 101*, 671-688.

Vansteenkiste, M., Zhou, M., Lens, W., & Soenens, B. (2005). Experiences of autonomy and control among Chinese learners: Vitalizing or immobilizing? *Journal of Educational Psychology, 97*, 468-483.

Vargas-Gomez, R. G. & Yager, R. E. (1987). Attitude of students in exemplary programs toward their science teacher. *Journal of Research in Science Teaching, 24*, 87-91.

Veermans, M., Lallimo, J., & Hakkarainen, K. (2005). Patterns of guidance in inquiry learning. *Journal of Interactive Learning Research, 16*, 179-194.

Verloop, N., Van Driel, J., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research, 35*, 441-461.

Vervoort, L. (2011). Leerkrachtstijl als onderdeel van een krachtige leeromgeving in het lager onderwijs: Een analyse van observatierapporten [Teacher style as part of a powerful learning environment in elementary school: An analysis of observation reports]. Unpublished Master's thesis, University of Leuven, Faculty of Psychology and Educational Sciences, Belgium.

Viilo, M., Seitamaa-Hakkarainen, P., & Hakkarainen, K. (2011). Supporting the technology-enhanced collaborative inquiry and design project: a teacher's reflections on practices. *Teachers and Teaching: theory and practice, 17,* 51-72.

Wallace, G. (1996). Engaging with learning. In J. Rudduck (Ed.), *School improvement: what can pupils tell us?* London: D avid Fulton.

Walma van der Molen, J. H. (2007). Eindrapport VTB Attitude Monitor. De ontwikkeling van een attitude-instrument op het gebied van wetenschap en techniek voor leerlingen in het basisonderwijs [Final rapport VTB attitude monitor. The development of an attitude instrument on the domain of S&T for pupils in primary education]. Den Haag, the Netherlands: Platform Bèta Techniek.

Walker, J.M.T. (2008). Looking at teacher practices through the lens of parenting style. *The Journal of Experimental Education, 76,* 218-240.

*Wang, J.-R., & Lin, S.-W. (2009). Evaluating elementary and secondary school science learning environments in Taiwan. *International Journal of Science Education, 31,* 853-872.

Watson, G. (2002). Using technology to promote success in PBL courses. *The Technology Source.* Retrieved from http://technologysource.org/article/using_technology_to_promote_success_in_pbl_courses/

Watters, J. J., & Ginns, I. S. (2000). Developing motivation to teach elementary science: Effect of collaborative and authentic learning practices in preservice education. *Journal of Science Teacher Education, 11,* 301-321.

Weld, J., & Funk, L. (2005). 'I'm not the science type': Effect of an inquiry biology content course on preservice elementary teachers' intentions about teaching science. *Journal of Science Teacher Education, 16,* 189-204.

Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality.* Princeton, NJ: Educational Testing Service. Retrieved from: www.ets.org/Media/Research/pdf/PICTEAMAT.pdf

Wenglinsky, H. (2002). The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10,* 1-30.

Wentzel, K.R. (2002). Are effective teachers like good parents? Teaching styles and student adjustment in early adolescence. *Child Development, 73,* 287-301.

Westwood, P. (2006). Teaching and learning difficulties: Cross-curricular perspectives. Victoria, Australia: ACER.

Wheatley, H. G. (1991). Constructivist perspectives on science and mathematics learning. *Science Education, 75,* 9-21.

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J.S. Eccles (Eds.), *Development of achievement motivation* (pp. 91-120). San Diego, CA: Academic Press.

Wilkins, J. L. M. (2008). The relationship among primary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education, 11,* 139-164.

Woodbury, S., & Gess-Newsome, J. (2002). Overcoming the paradox of change without difference: A model of change in the arena of fundamental school reform. *Educational Policy, 16,* 763-783.

Wubbels, T., & Brekelmans, M. (2012). Teacher-students relationships in the classroom. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (Chapter 80, pp. 1241-1255). Dordrecht: Springer. Retrieved from http://link.springer.com/chapter/10.1007%2F978-1-4020-9041-7_80

Wetzels, A. (2015). *Curious minds in the classroom: The influence of video feedback coaching for teachers in science and technology lessons.* Unpublished doctoral dissertation, University of Groningen, Faculty of Behavioural and Social Sciences, The Netherlands.

Windschitl, M. (2003). Inquiry projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Education, 87,* 112-143.

Wurdinger, S., & Haar, J., Hugg, R., & Bezon, J. (2007). A qualitative study using project-based learning in a mainstream middle school. *Improving Schools, 10,* 150-161.

Yates, S., & Goodrum, D. (1990). How confident are primary school teachers in teaching science? *Research in Science Education, 19,* 183-204.

Yerrick, R.K. (2000). Lower track students' argumentation and open inquiry instruction. *Journal of Research in Science Teaching, 37,* 807-838.

Yerrick, R., Parke, H., & Nugent, J. (1996). Struggling to promote deeply rooted change: The "filtering effect" of teachers' beliefs on understanding transformational views of teaching science. *Science Education, 81,* 137-159.

*Yilmaz, H., Turkmen, H., Pedersen, J. E., & Huyuguzel Cavas, P. (2007). Evaluation of pre-service teachers' images of science teaching in Turkey. *Asia-Pacific Forum on Science Learning and Teaching,* 8.

Young, T. (1998). Student teachers' attitudes towards science (STATS). *Evaluation and Research in Education, 12,* 96-111.

Young, T. (1998). Student teachers' attitudes towards science (STATS). *Evaluation and Research in Education, 12,* 96-111.

*Young, B. J., & Lee, S. K. (2005). The effects of a kit-based science curriculum and intensive science professional development on elementary student science achievement. *Journal of Science Education and Technology, 14,* 471-481.

Zembal-Saul, C., Krajcik, J., & Blumenfeld, P. C. (2002). Elementary student teachers' science content representations. *Journal of Research in Science Teaching, 39,* 443-463.

Zion, M., & Slezak, M. (2005). It takes two to tango: In dynamic inquiry, the self-directed student acts in association with the facilitating teacher. *Teaching and Teacher Education: An International Journal of Research and Studies, 21*, 875-894.

List of tables

Study 4

Dankwoord

Dit PhD onderzoek is het resultaat van vier verrijkende jaren aan de KU Leuven. In deze periode kon ik verscheidene malen rekenen op de hulp en ondersteuning van mijn promotor, co-promotoren, collega's, vrienden en familie. Zonder hen was dit proefschrift niet mogelijk geweest. Ik ben dan ook heel dankbaar voor de ondersteuning die ze tijdens dit doctoraatstraject gegeven hebben. In de eerste plaats wil ik mijn promotor, Ferre Laevers, en copromotoren Lieven Verschaffel en Wim Van Dooren, bedanken voor hun begeleiding in de verschillende fases van het traject. Door Ferre' s inspirerende visie op onderwijs en praktijkgericht onderzoek voelde ik me gesterkt om het onderzoek de richting te geven die het nu is uitgegaan. De begeleiding van Lieven Verschaffel en Wim Van Dooren als co-promoteren was essentieel in dit traject. Hen wil ik bedanken voor hun toegewijde inzet om van bij de start het design mee uit te zetten en telkens opnieuw gedetailleerd feedback te geven op mijn teksten. In het bijzonder dank ik Lieven voor zijn fijnmazige blik als het gaat om het helder structureren en eenduidig gebruik van concepten. Wim, de helderheid die jij steeds weer in complexe kwesties, gaande van het design tot de dataverwerking- en analyses, wist te brengen, brachten me tot dieper inzicht en relativering.

Tijdens de voorbije vier jaren kon ik ook rekenen op de steun van vele enthousiaste collega's. In het bijzonder wil ik Evelien Buyse bedanken voor haar onuitputtelijke ondersteuning bij de dataverwerking. Met haar uitvoerige statistische expertise toonde zij zich een coach die altijd bereid was om mijn vragen uitvoerig te beantwoorden. Een bijzonder woord van dank gaat uit naar Thaline Stas voor haar volgehouden inzet bij het scoren van de leerlingbetrokkenheid voor dit PhD onderzoek. CEGO-collegas' Charlotte Van Cleynenbreugel, Nathalie Michalek, Marian De Groof, Hanne Vanmierlo, Mieke Daems, Inge Laenen, Veronique Steen, Joanna Papieska, Lies Huysmans, Sofie Lietaert en Bart Declercq wil ik bedanken voor hun fijne babbels en interesse in mijn werk. De gedeelde visie op onderwijs en opvoeding en de gezellige lunchpauzes en teamdagen droegen bij tot een collegiale sfeer op CEGO. Voor al deze momenten dank ik ook Els Menu, Ilse Aerden, Daisy Boonen, Heidi Defever, Inne Jackers, Inneke Berghmans, Marta Klejman, Laura Eeckeleers, Thaila Stoffelen, Aäron Vaneeckhout, Griet De Bruyckere, Veerle De Winter, Ilse Aerden, Sarah Currinckx en Ludo Heylen. Diverse collega's van het TalentenKracht (TK) consortium, met name Sabine Van Vondel, Carla Geveke, Astrid Menninga, Annemie Wetzels, Henderien Steenbeek, Eliane Segers, Joep van der Graaf, Barbara Wagensveld, Tessa van Schijndel en Willemijn Schot, wil ik bedanken voor de fijne samenwerking tijdens verschillende TK bijeenkomsten en symposia op de EARLI conferenties.

Alle scholen, directies, leerkrachten en leerlingen die aan dit onderzoek hebben meegewerkt, wil ik bedanken voor hun bereidheid tot participeren en hun betrokkenheid doorheen het hele onderzoekstraject. De MES-studenten Lucy Hunt en Mary Noll wil ik hartelijk danken om op korte tijd mijn teksten na te lezen op het Engels.

Ook vele vrienden ben ik dankbaar voor hun luisterende oor de afgelopen jaren. De etentjes en wandelingen met Wing, Hilke, Christine, Jennifer, Steffi en Heidi maakten dat de tijd om te ontspannen aangenaam werd ingevuld. Ook de relaxte samenkomsten met Tim, Veerle, Hans, Hanne, Brecht, Isa, Choi Yoo-Hyun en Laura-An, waren een welkome afwisseling bij het schrijven.

Een grote dankbaarheid gaat uit naar mijn ouders, Hilde en Ronald, mijn broer Steven en zijn partner Stephanie, en mijn oma's Denise en Annetta, die mij steeds aangemoedigd hebben bij het ontplooien van mijn interesses – en dus ook van dit doctoraat. Steven wil ik in het bijzonder bedanken voor het verzorgen van de lay-out van dit proefschrift. Ik wil ook Hans, Katelijne, Jasper, Tine en René bedanken voor de warme welkom en de hartelijke babbels op meerdere zondagen. Mijn laatste woord van dank gaat uit naar Joris, die dit proces van zeer nabij gevolgd heeft. Woorden schieten te kort om hem te bedanken. Zijn liefde, empathie en de aanmoediging die hij elke dag opnieuw gaf, hebben me er doen in slagen dit doctoraat te schrijven.